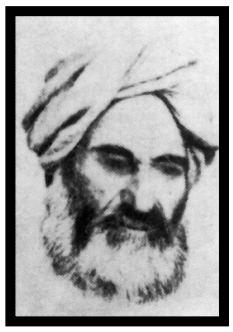

JOURNAL OF SCIENCES

University of Tehran

Abu Reyhan al-Biruni
(973-1048)
Iranian Scientist

ISLAMIC REPUBLIC OF IRAN

ISSN 1016-1104

Vol. 36, No. 1, Winter 2025

CONTENTS

بسم الله الرحمن الرحیم

In the Name of Allah, the Beneficent, the Merciful

JOURNAL OF SCIENCES

ISLAMIC REPUBLIC OF IRAN

Journal of Sciences, Islamic Republic of Iran is published quarterly by The University of Tehran,
16th St., North Kargar Ave, Tehran, Islamic Republic of Iran
Tel.: (0098-21)88334188, 88012080-212
Fax: (0098-21) 88334188
P.O. Box 13145-478
http://jsciences.ut.ac.ir
E-mail: jsciences@ut.ac.ir
E-mail: nooridaloii@sina.tums.ac.ir

ISSN 1016-1104

Instructions to Authors

1. General Policy

The *Journal of Sciences, Islamic Republic of Iran* (*J. Sci. I. R. Iran*) is published quarterly by the University of Tehran. Contributions from all fields of basic sciences may be submitted by scientists from all over the world.

The papers submitted to this journal should not have been published previously, except in the form of a brief preliminary communication, nor submitted to another journal. The decision to accept a contribution rests with the Editorial Committee of the *J. Sci. I. R. Iran.* Manuscripts will be considered for publication in the form of articles, preliminary communications, notes and review articles. The work should be original or a through review by an authoritative person in a pertinent field.

2. Copyright

Submission of a manuscript implies that the author(s) agree to transfer copyright to the *J. Sci. I. R. Iran* when the contribution is accepted for publication. Reproduction of the text, figures, or tables of this journal is allowed only by permission of the Editorial Committee.

3. Preparation of Manuscripts

3.1. General considerations. Manuscripts must be submitted in English according to Journal Instructions (It is necessary to submit at least 2 files including "Title page" and "Main file"). They **must be** typewritten in **Microsoft Word** (all versions). Authors are requested to reserve margins of at least 3 cm at the top and bottom of each page and at least 4 cm on the left-hand side.

Tables and illustrations (both numbered in Arabic numerals) should be prepared on separate pages. Tables require a heading and figures a legend, also prepared on a separate page. In Electronic submission, figures should be with the following caveats: all figures should be submitted at a minimum of 300 dpi and saved as TIFF files (avoid submitting JPEG files) after your text file.

Manuscripts should be kept to a minimum length and should be subdivided into labeled sections (**Introduction**, **Materials and Methods**, **Results**, **Discussion**, **Acknowledgement**, **References**). A current issue of *J. Sci. I. R. Iran* should be consulted.

3.2. ***Title page.*** The title of a manuscript should reflect concisely the purpose and findings of the work in order to provide maximal information for a computerized title search. Abbreviations, symbols, chemical formulae, references, and footnotes should be avoided.

The authors' full first names, middle initials and last names should be given, followed by the address(es) of the contributing department(s). (e.g. Department, Faculty, University, City, Country).

Telephone, Fax and Email of corresponding author should be footnoted on the bottom of the first page of each manuscript.Footnotes may be added to indicate the present mailing address(es) of the author(s). (e.g. *Corresponding author, Tel: 00982188012080, Fax: 00982188012081, E-mail: jsciences@ut.ac.ir).

Special types of print should be indicated as follows:

Type	Mark	Symbol	Example	Example (Printed)
Boldface[a]	Single underline		Introduction	**Introduction**
Italic[b]	Wave-like underline		In vivo	*In vivo*
Small capital[c]	Double underline		0.2 m	0.2 M
Boldface italic[d]	Underline with wavy line		R	***R***

a) Headings, designated numbers of chemical compounds, subheadings.

b) Configurational prefixes ((R)-, (S)-, cis-, trans-, tert-, etc); Latin words or abbreviations, words in languages other than English; trade names of compounds (first letter should be capitalized); names of authors if mentioned in the text.

c) Symbols of molar and normal concentrations (M and N), D- and L-, the names or initials of the nomenclature of species.

d) Italicized terms and prefixes in headings.

The total number of pages (including references, tables, copies of formula collections (if any), schemes and figures should be marked in the upper left-hand corner of the first page of each copy. The complete address, including phone number, and E-mail address of the correspondence author should also be given.

3.3. ***Main File:***

a) The title of the article (The first letter of each word must be capital).

b) Abstract should be self-explanatory and intelligible without references to the text and titles, it must not exceed 250 words.

c) At least between three to five keywords should be chosen by the author(s).

d) Introduction

e) Materials and Methods

f) Results and Discussion (they can be separate section too)

g) Figures and Tables (if There are any)

Theoretical articles must have atleast two main title : Introduction and Results.

3.4. References. References may be numbered alphabetically or sequentially in the order they are cited in the text. References typed with double spacing are to be listed in numerical order at the end of the main text. They should be addressed according to the following examples:

Journals:

Noori-Daloii M.R., Swift R.A., Kung H.J., Crittenden L.B., and Witter R.L. Specific integration of REV proviruses in avian bursal lymphomas. *Nature*, **294** (5841): 574-576 (1981).

Noori Daloii M.R., Saffari M., Raoofian R., Yekaninejad M., Saydi Dinehkabodi O., and Noori-Daloii A.R. The multidrug resistance pumps are inhibited by silibinin and apoptosis induced in K562 and KCL22 leukemia cell lines. *Leukemia Research*, **38** (5): 575-580 (2014).

Books:

Rang G.M. and Petrocelli S.R. *Fundamentals of Aquatic Toxicology.* Hemisphere Publishing Corporation, New York, 1129 p. (1991).

Chapters in Books:

Walsh J.H. Gastrointestinal hormones. In: Johnson L., Christensen J., Jackson M., Jacobson E., and Walsh J.H. (Eds.), *Physiology of the Gastrointestinal Tract*, 2nd Ed., Raven, New York, pp. 181-254 (1987).

Thesis:

Kossir A. Extraction liquid-liquid du Zinc (II) en milieu cyanure. Application ala valorization des minerais de zinc oxides, Ph.D. *Thesis*, University of Paris (VI), 116 p. (1991).

Please note that papers with incorrect formatted references will be returned.

4. Forwarding Address

Manuscripts should be submitted to the journal site (http://jscinces.ut.ac.ir).

M.R. Noori-Daloii, Ph.D.
J. Sci. I. R. Iran
University of Tehran
P.O. Box 13145-478 Tehran, I. R. Iran
http://jsciences.ut.ac.ir
E-mail: jsciences@ut.ac.ir

Journal of Sciences, Islamic Republic of Iran 36(1): 5 - 19 (2025) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Genetic Analysis of Y-STRs in Two Iranian Sub-Populations

M. Ghaderi-Zefrehei[1], F. Rafeie[2], A. Alipour Tabrizi[3], Z. Baratieh[4], S. Kamal Fotouhi[4], J. Smith[5], M. Muhaghegh Dolatabady[1*]

[1] *Department of Animal Science, Faculty of Agricultural Sciences, Yasouj University, Yasouj, Islamic Republic of Iran*
[2] *Department of Agricultural Biotechnology, Faculty of Agricultural Sciences, University of Guilan, Rasht, Islamic Republic of Iran*
[3] *Molecular Genetics Department, Mashhad Forensic Medical Center, Mashhad, Islamic Republic of Iran*
[4] *Forensic Medical Research Center, National Forensic Medical Organization, Tehran, Islamic Republic of Iran*
[5] *The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK*

Received: 3 December 2024 / Revised: 7 October 2025 / Accepted: 8 November 2025

Abstract

This study presents a comprehensive genetic analysis of 17 Y-chromosomal short tandem repeat (Y-STR) loci in two Iranian sub-populations from the Fars (n=109) and Isfahan (n=180) provinces. The loci investigated included DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 (Y-GATA-C4), and Y-GATA-H4. Results demonstrated that the DYS385a/b locus exhibited the greatest allelic diversity in both populations, with 11 distinct alleles detected and mean allele counts of 6.29 and 5.88 in the Fars and Isfahan groups, respectively. Conversely, the Fars cohort showed the lowest allelic variation (three alleles) at DYS439 and Y-GATA-H4 loci, while the Isfahan population exhibited minimal variation (four alleles) at DYS19 and DYS439. Haplotype analyses revealed intra-population sharing rates of 2.75% in Fars and 10.0% in Isfahan, with an overall 8.3% haplotype overlap observed across the combined dataset of 289 individuals. Both populations exhibited high haplotype diversity values approaching 0.99, indicating substantial genetic variability. The haplotype discrimination capacity varied among populations, with value of 0.9725 for Fars, 0.8519 for Isfahan, and 0.9170 for the entire sample set. Population differentiation was assessed using pairwise FST and RST metrics, which confirmed significant genetic divergence between Fars and Isfahan groups (FST = 0.00743, $p < 0.001$; RST = 0.0106, $p < 0.01$). These findings underscore the genetic distinctness of the two sub-populations. The study highlights the necessity for further research incorporating Y-chromosomal single-nucleotide polymorphisms (Y-SNPs), larger sample sizes, and additional ancestral information to enhance the understanding of genetic structure and demographic history within Iranian populations.

Keywords: Haplotype; Genetic diversity; Y-STR markers.

* Corresponding Author: Tel: +98 74331006000; Email: mmuhaghegh@yu.ac.ir

Introduction

The genome of most organisms contains repetitive DNA sequences that are very small in length (2 to 7 bp), known as microsatellites, or commonly called short tandem repeats (STRs) (1). The STRs have become popular among repetitive DNA markers because of high variability in their repeat numbers and easy determination of their genotypes (2). The STR loci, as a genetic marker, have been widely used for individual identification (3, 4), population characterization (5), molecular diagnosis of genetic diseases (6), and even forensic studies (7). The STRs on the Y chromosome and its associated Y-STR haplotype, makes it potentially valuable for forensic genetic particularly in sexual assault investigations (8). Males commit the vast majority of violent crimes; therefore, a crime-scene sample left by a culprit will usually be informative when typed with Y chromosome-specific markers (9), and even if no Y-specific markers were found in a sample, this result would still be helpful. In addition, Y-STR analysis readily provides an assailant-specific profile in cases of Azoospermia (10).

Iran is a heterogeneous country that makes up about 1.12% of the total world population (https://www.worldometers.info/) has significant diversity due to factors such as climate, agriculture, industry, ethnicity, religion and language (11). These factors lead to significant genetic diversity, although the Iranian population is studied poorly, particularly concerning Y-STR data. Mixed signals of gene diversity are observed in Iranian populations. For example, no significant difference has been observed in genetic distance between the Iranian Gilak, Mazandarani, Bakhtiari, and non-Iranian Turkish, Azerbaijanian, Armenian, and Kurdish ethnic groups using Y-STRs. However, a significant gene distance was observed among Talysh, the Iranian Arabs, Georgian and Kazakh ethnic groups (12). A total of 204 unique Y-STR haplotypes were detected in unrelated male samples from Mazandaran and Gilan provinces. However, no unique Y-STR haplotypes were observed across the two provinces (13). The results of a study showed the efficiency of Y-STRs as a reliable tool in forensic genetic medicine and the evolutionary comparison of different ethnic groups from northern Iran (14). In addition, in another study, based on whole genome data from eleven ethnic groups present in Iran, it has been observed that the Iranian population consists of genetically clusters of overlapping ethnic groups, several strongly admixed groups, and shows distinct genetic diversity with respect to populations in close geographic proximity (15). In the current study, the Y-STR profiles were investigated in two Iranian sub-populations (Isfahan and Fars provinces). These two regions cover a large section of the ancient capitals of the Persian Empire, in which population genetics profile analysis has rarely been undertaken. The study had two primary aims. First, forensics, to provide a database with high-quality haplotypes with more reliable Y-STRs markers than currently available studies which have been conducted in these regions. Second, anthropological, as the data and results can potentially be reused by any human population anthropologist to reconstruct the history of the Iranian or surrounding populations.

Materials and Methods

Populations and DNA profiling

Through confirmation of ethnic identity (due to the mixed nature of the two populations via continual migration), only healthy and unrelated male participants capable of tracing their ancestry for at least two generations were included. The target sample sizes of n = 109 for Fars and n = 180 for Isfahan were determined a priori using G*Power 3.1 software to ensure ≥80% statistical power ($\alpha = 0.05$) to detect a medium effect size ($f = 0.25$) in haplotype diversity differences between the two populations, based on previously reported genetic variation in Iranian subpopulations (13). This calculation ensured sufficient sensitivity to detect meaningful population differentiation while accounting for the expected within-population heterogeneity inherent in Y-STR markers. Y-DNA samples from two socio-geographical populations in the Fars (109 samples) and Isfahan (180 samples) provinces were collected by the Iranian Medical Research Council and the Iranian Forensic Genetics Bank, in collaboration with two referral centers in the capital cities of the two provinces. DNA was isolated from bloodstains on 1.2 mm FTA cards using the QIAamp DNA Investigation Kit (Qiagen). Multiplex polymerase chain reaction (PCR) was performed to amplify the 17 Y-STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390,

DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635/Y-GATA-C4, and Y-GATA-H4) using the commercially available AmpFLSTR® Yfiler™ PCR Amplification Kit (Applied Biosystems, Foster City, CA). This kit employs a pre-designed, multiplexed primer set optimized for human Y-chromosome STR analysis; the primer sequences are proprietary and not publicly disclosed by the manufacturer. However, the amplicon sizes, fluorescent labeling (6-FAM, NED, PET, VIC), and amplification conditions are standardized and validated according to the manufacturer's specifications (Applied Biosystems). All reactions were carried out under the following thermal cycling parameters: initial denaturation at 95°C for 11 min, followed by 28 cycles of 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min, with a final extension at 60°C for 45 min. Multiplex PCR products were analyzed on an automated ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) using POP-4 polymer in a 36 cm capillary array. Quality control for Y-STR typing was confirmed by successful genotyping of five blind reference samples provided by the Y-Chromosome Haplotype Reference Database (YHRD; https://yhrd.org/) (16). Haplotypes were interpreted using GeneMapper ID-X Software v1.3 (Applied Biosystems), and allele calling was based on the Nei equation (17). The obtained Y-STR data were submitted to the YHRD under Accession Numbers YA004229 (Fars) and YA004228 (Isfahan). The general formal ethical approval for this study was obtained from the Ethical Committee of the Legal Medicine Organization, Tehran (Reference ID: 1396.0018).

Y-STR data analyses

GenAlEx V6.5 (18) was employed to determine allelic frequencies, number of alleles, number of matches between haplotypes, and Shannon index for the 17 Y-STR markers. The Shannon's Index is an information statistic index, assuming all types are represented in a sample and that they are randomly sampled. It combines both evenness and richness in a single measure (19). Gene diversity was also calculated based on the Nei equation (17). Also, shared haplotypes, theta parameters (θ; the number of mutations per generation, assuming that the population is in mutation-drift equilibrium), and expected heterozygosity (HE) were estimated using Arlequin software (20). Theta was calculated from the expected homozygosity (HomE) as follows:

$$\theta = \frac{1\text{-HomE}}{\text{HomE}}$$

where, HomE = 1 − HE. Moreover, haplotype discrimination capacity (DC) was calculated as follows (21):

$$DC = \frac{\text{Number of unique haplotypes}}{\text{Number of total genotype markers}}$$

Additionally, the probability of exclusion (PE) was estimated by (22):

$$PE = 1\text{-} \sum P_i^2$$

where P is the haplotype frequency.

HapYDive software (http://www.portugene.com/hapydive.html) (23) was also used to calculate the number of different haplotypes and haplotype diversity (HD). The haplotype match probability (HMP) was calculated as $HMP = 1 - HD$. Haplogroup Predictor software (www.hprg.com/hapest5/hapest5a/hapest5.htm) was used to predict the haplogroup from the haplotypes (24,25). The 17-marker haplotypes were analyzed using the following 26 haplogroups (as defined by ISOGG 2017 Y Tree): C2-M217, D-M174, E1a-M132, E1b1a-V38, E1b1b-M35, G1-M342, G2a-P15, G2b-M377, H-M69, I1-M253, I2a-P37(xM26), I2a-M26, I2b-M223(xM284), I2b-M284, J1-M267, J2a-M410, J2b-M12, L-M11, N-M46, O1-F265, O2-M122, Q-M242, R1a-M17, R1b-M349, R2-M124, and T-M184. Generally, a probability of 85% with a fitness score of 20 was considered acceptable for prediction. In a few cases, a fractional result was obtained for STRs, but since the Haplogroup Predictor software cannot handle such values, the values were truncated before being analyzed. Analysis of Molecular Variance (AMOVA) and of F_{st} and R_{st} values were calculated using GenAlEx V6.5 and YHRD (Release 59- 2018/Nov/01) softwares (p-values computed by 10,000 permutations) for obtaining the 17 Y-STR haplotypes and also data from neighboring countries, including Pakistan (Pathan), Turkey, and Iraq (Table S1).

Results

Despite the existence of manuscripts with larger sample sizes exploring Iranian subpopulations, this paper stands out for explicitly focusing on two key Iranian capital regions that were once capitals of the Old Persian Empire. These regions cover a significant portion of the Iranian plateau. Furthermore, we haven't only conducted a forensic genetic analysis but also delved into demographic analysis using Y-STR data.

In the current study, the Y-STR profiles were analyzed in two Iranian subpopulations (Isfahan and Fars provinces), including 289 random, unrelated forensic samples. Haploid allelic frequencies and number of alleles for the 17 Y-STR loci in Fars and Isfahan populations are shown in Figure 1 and Table S2. The results of estimated genetic diversity parameters are shown in Figure 2a, b, c and d and Table S3. The

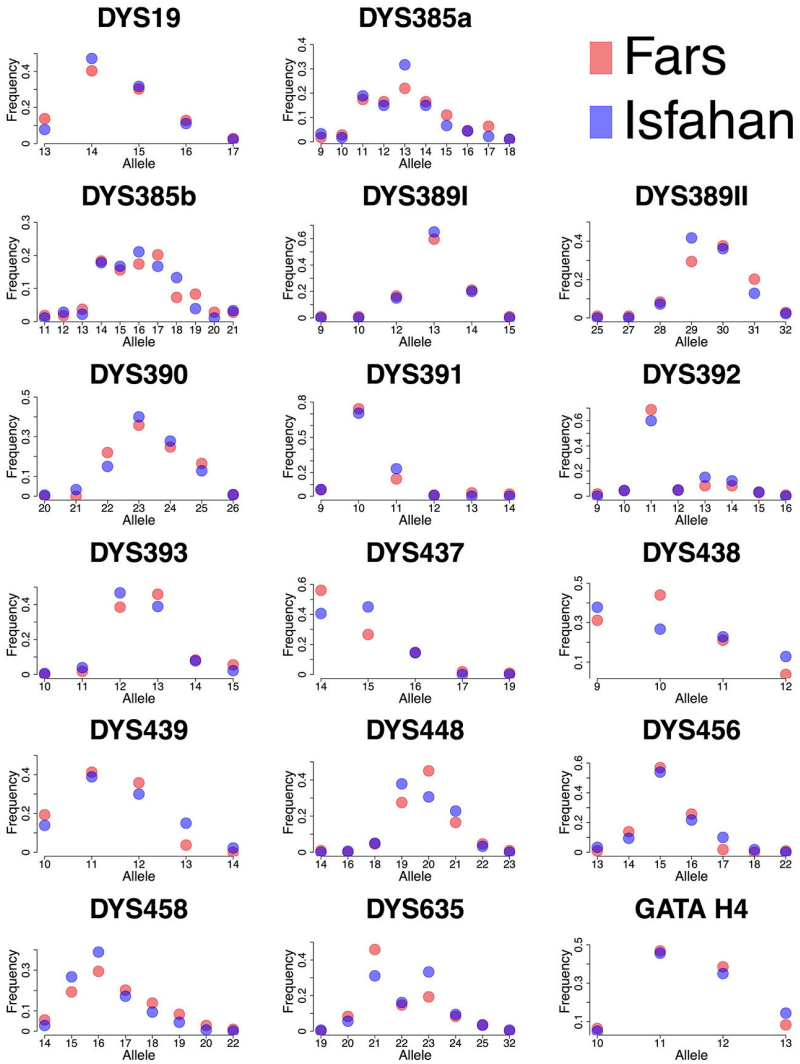

Figure 1. Haploid allelic frequencies of 17 STRs in Fars and Isfahan populations are presented in red and blue, respectively (purple represents overlap). For detailed information about each population, see Tables S1 and S2.

maximum number of alleles (11 alleles) was found in the DYS385a/b locus in both populations. The minimum number of alleles was observed for the DYS439 and GATA-H4 loci, 3 and 4 alleles each in Fars and Isfahan, respectively (Figure 2a). For further analysis of the DYS385a/b locus, the Shannon index values for Fars (2.092) and Isfahan (2.025) populations and the HE values (0.86 for Fars and 0.85 for Isfahan) were calculated (Figure 2b and c, and Table S3). In samples from the Fars province, the highest values for the HE parameter were detected in DYS385b (0.86), DYS385a (0.85), DYS458 (0.81), DYS390 (0.74), and DYS389II (0.73). Similarly, these values for Isfahan samples were in DYS385b (0.85), DYS385a (0.81), DYS635 (0.75), DYS390 (0.74), and DYS458 (0.72). The highest values of Theta (θ) for DYS385b were 26 in Fars and 22 in Isfahan individuals, followed by 23 (DYS385a) and 13 (DYS458) in Fars and 14 (DYS385a) and 7 (DYS635) in Isfahan samples (Figure. 2d). Pairwise population differences were based on the distance method and were used to calculate F_{st} values. Even though a comparative analysis of the Isfahan population indicated that the population is slightly different from the Fars population (Fst = 0.00743, $p<0.001$), it is hardly can be said it is statistically different as its values obtained are low and less than that 0.15. The R_{st} value obtained for the Isfahan

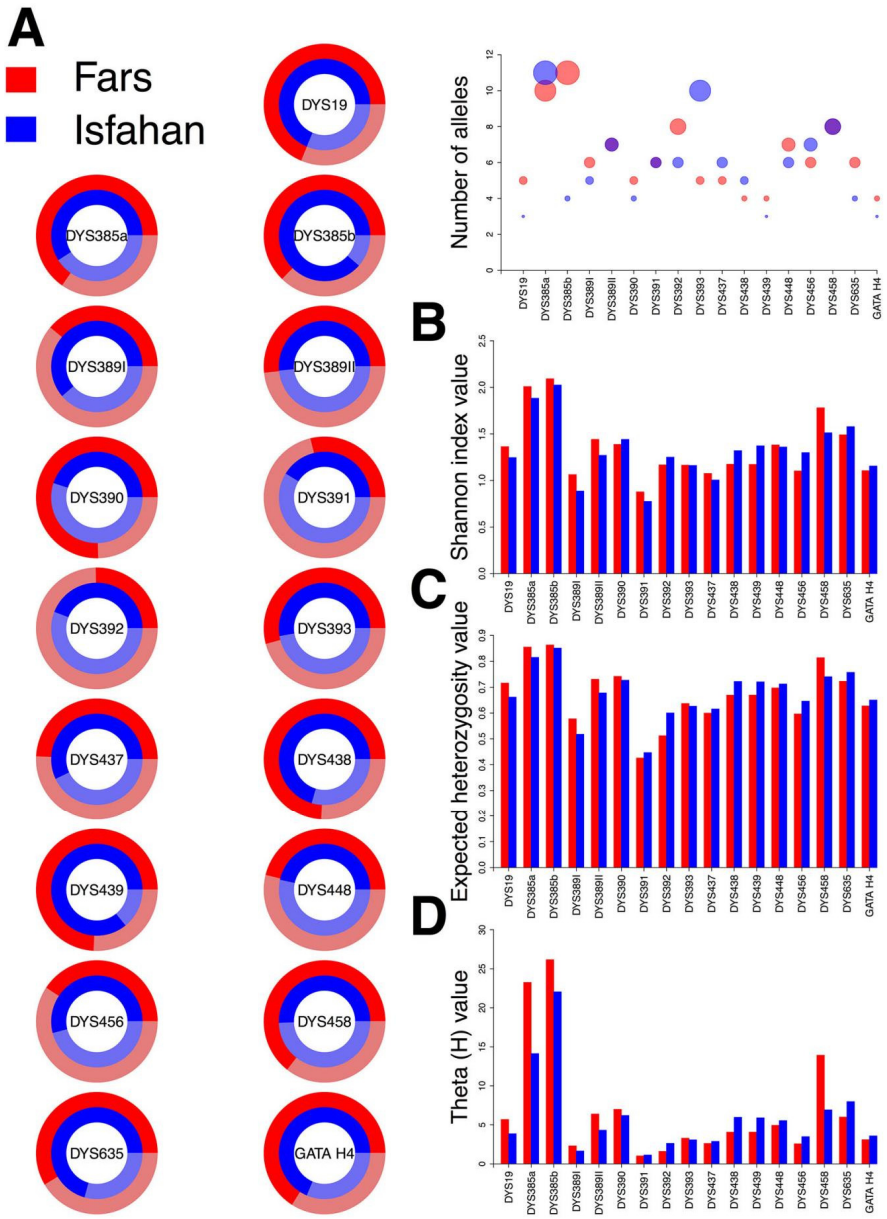

Figure 2. Number of alleles, gene diversity and theta values of 17 STRs in Fars (red) and Isfahan (blue) populations, Iran (purple represents overlaps). For more information about each population, see Table S3. A: Daunt charts show total number of alleles in each STR where the effective number of alleles is shown in bold and non-effective are shown in pale colours. Circle size of the corresponding plot represents number of alleles for each STR of the studied populations. DYS389II and DYS391 had the exact same number of alleles in both Fars and Isfahan populations (7 and 6, respectively). B: Shanon index, C: Expected heterozygosity and D: Theta (θ).

and Fars provinces was 0.0.0106 (p<0.01). The numbers of total and unique haplotypes in each studied population, along with the number of haplotypes shared within and between the Fars and Isfahan populations, are shown in Table 1. The shared haplotypes were either cases of two or three matching haplotypes. There were 106 distinct haplotypes within the Fars population, while three of the haplotypes occurred twice. There were 161 distinct haplotypes within the Isfahan population while 3 and 16 haplotypes occurred three and twice times, respectively.

Table 1. Haplotype and forensics parameters for 17 Y-STR loci deriving from 289 samples from Isfahan and Fars provinces of Iran

Parameters	Fars	Isfahan	Total Population
Number of haplotypes	109	189	289
Number of Different haplotypes	106	161	265*
Shared haplotypes			
Cases of 3 matching haplotypes	0	3 (1.58%)	3 (1.03%)
Cases of 2 matching haplotypes	3 (2.75%)	16 (8.46 %)	19 (6.57%)
Total shared haplotypes	3 (2.75%)	19 (10.05%)	24 (8.30%)
Forensic parameters			
Haplotype discrimination capacity (DC)	0.9725	0.8519	0.9170
Chance of exclusion (CE)	0.6676	0.6720	0.6698
Haplotype diversity (HD)	0.9995	0.9984	0.9989
Gene diversity	0.6737	0.6758	0.6774
Haplotype match probability (HMP)	0.0005	0.0016	0.0011

* Two haplotypes were shared between Fars and Isfahan populations, so the number in total was reduced by 2

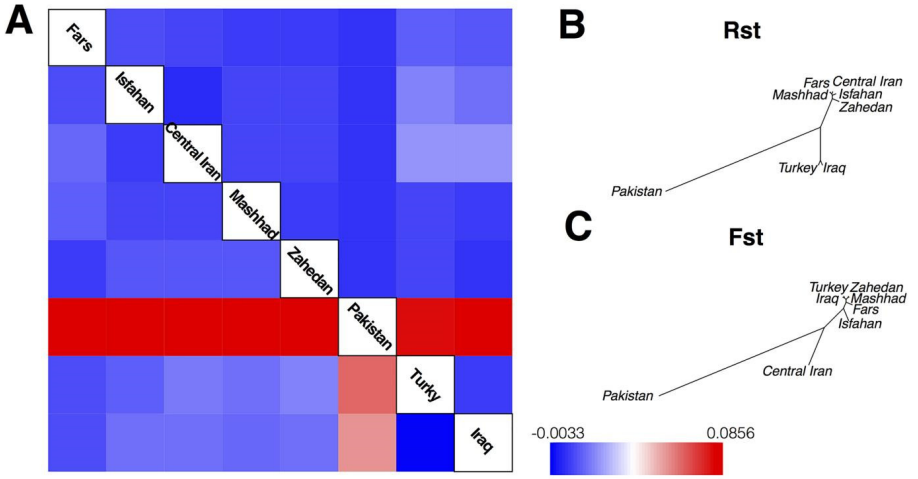

Figure 3. The AMOVA analysis (F_{st} and R_{st} values) of the Y-Chromosome Haplotype Reference (YHRD (https://yhrd.org/) (3); *P*-values computed by 10,000 permutations). The off-diagonal entries in the matrix provide the F_{st} and R_{st} values between each variable pair. For more information about each population see Table S1. A: The values below the diagonal are R_{st} and values above the diagonal are F_{st} obtained for comparison between five Iranian and three regional populations (Tables S4 and S5), B: The unrooted tree of R_{st} values of five Iranian and three regional populations (Table S4) C: The unrooted tree of F_{st} values obtained for comparison between five Iranian and three regional populations (Table S5).

Two haplotypes from Fars matched with two haplotypes from Isfahan, but did not match with any other haplotypes within the populations. In total, 24 (8.30%) of the 289 haplotypes were shared between and within subpopulations. The population genetics parameters DC, CE, HD, gene diversity, and HMP were also analyzed (Table 1). Different populations from Iran and neighboring countries were used for regional analysis (Table S1). The highest and lowest R_{st} values were seen for Pakistan-Fars (0.0856) and Turkey-Iraq (-0.0008), respectively (Figure 3a and b and Table S4). The highest and lowest F_{st} values were observed for Central Iran-Isfahan (-0.0033) and Pakistan-Central Iran (0.0056), respectively, as shown in Figure 3a and c and Table S5. Allelic frequencies can be different between different Y haplogroups. Indeed, those differences form the basis for haplogroup prediction from Y-STR values. Table 2 shows the predicted haplogroup frequencies for the Fars and Isfahan samples, plus the summary for the combined populations. Note that the SNP names in Table 2 are only for haplogroup identification purposes - no SNPs were tested in the current study. As can be seen in Table 2, the highest haplogroup frequency was observed for J2a and R1a haplogroups. Unfortunately, in 12 out of 280 haplotypes, no haplogroup prediction could be made with the set of haplogroups employed in the prediction

Table 2. Haplogroup Summary Statistics

All Samples			Fars Samples			Isfahan Samples		
Haplogroup	Count	Percent	Haplo	Count	Percent	Haplo	Count	Percent
C2-M217	2	0.7%	C2-M217	2	1.8%	C2-M217	0	0.0%
D-M174	1	0.3%	D-M174	1	0.9%	D-M174	0	0.0%
E1a-M132	0	0.0%	E1a-M132	0	0.0%	E1a-M132	0	0.0%
E1b1a-V38	4	1.4%	E1b1a-V38	0	0.0%	E1b1a-V38	4	2.2%
E1b1b-M35	23	8.0%	E1b1b-M35	13	11.9%	E1b1b-M35	10	5.6%
G1-M342	6	2.1%	G1-M342	2	1.8%	G1-M342	4	2.2%
G2a-P15	15	5.2%	G2a-P15	9	8.3%	G2a-P15	6	3.3%
G2b-M377	0	0.0%	G2b-M377	0	0.0%	G2b-M377	0	0.0%
H-M69	10	3.5%	H-M69	5	4.6%	H-M69	5	2.8%
I1-M253	1	0.3%	I1-M253	1	0.9%	I1-M253	0	0.0%
I2a-P37(xM26)	3	1.0%	I2a-P37	2	1.8%	I2a-P37	1	0.6%
I2a-M26	0	0.0%	I2a-M26	0	0.0%	I2a-M26	0	0.0%
I2b-M223(xM284)	0	0.0%	I2b-M223	0	0.0%	I2b-M223	0	0.0%
I2b-M284	1	0.3%	I2b-M284	0	0.0%	I2b-M284	1	0.6%
J1-M267	15	5.2%	J1-M267	10	9.2%	J1-M267	5	2.8%
J2a-M410	74	25.6%	J2a-M410	20	18.3%	J2a-M410	54	30.0%
J2b-M12	5	1.7%	J2b-M12	0	0.0%	J2b-M12	5	2.8%
L-M11	19	6.6%	L-M11	4	3.7%	L-M11	15	8.3%
N-M46	3	1.0%	N-M46	2	1.8%	N-M46	1	0.6%
O1-F265	1	0.3%	O1-F265	0	0.0%	O1-F265	1	0.6%
O2-M122	2	0.7%	O2-M122	1	0.9%	O2-M122	1	0.6%
Q-M242	11	3.8%	Q-M242	5	4.6%	Q-M242	6	3.3%
R1a-M17	37	12.8%	R1a-M17	16	14.7%	R1a-M17	21	11.7%
R1b-M343	22	7.6%	R1b-M343	3	2.8%	R1b-M343	19	10.6%
R2-M124	12	4.2%	R2-M124	3	2.8%	R2-M124	9	5.0%
T-M184	10	3.5%	T-M184	6	5.5%	T-M184	4	2.2%
Unknown	12	4.3%	Unknown	4	3.7%	Unknown	8	4.4%
Total	289		Total	109		Total	180	

program. This is probably due to four main factors: 1) the relatively small number of Y-STR markers (17 Y-STRs), 2) the omission in the set of 17 Y-STRs of markers of particular importance to haplogroup prediction (DYS388, YCAII), 3) the likelihood that the haplotypes in these studied populations have diverged somewhat from the populations represented in the predictor program database, and 4) the possible existence in these populations of haplogroups that are not included in the analysis. A total of 22 haplogroups were found in the two populations. About a quarter of the total population was found to be in the J2a-M410 haplogroup, while the next most common haplogroup was Haplogroup R1a-M17 at about 13%. In addition, 7.3% of the combined populations were found to be in haplogroup R1b-M343, 7.3% in E1b1b-M35, 7.3% in J1-M267, 6.7% in L-M11, 4.7% in G2a-P15, 4.2% in R2-M124, 3.9% in Q-M242, and 2.8% in H1-M69. The remaining haplogroups occurred at frequencies of 2% or less.

Discussion

Different ethnic groups, religious affiliations, languages, climates, *etc.,* have produced a heterogeneous population in Iran. Thus, exploring Y-STRs in Iran is a compelling venture. Both F_{st} (the classical allele frequency-based differentiation estimator) and R_{st} (the mutational based model differentiation estimator) measures, though significant, showed low values in this study, indicating that genetic variance of within-population is much larger than between population. Many Y-STRs-based studies have reported both F_{st} and R_{st}. For example, the most significant genetic distance (R_{st}=0.35434) was observed in Gilaks and Azeris; but, Sistanis and Lurs groups showed the smallest genetic distance (R_{st}=0.00483) compared to other Iranian ethnicities (26). However, there is a longstanding debate about using both measures in terms of their variance sampling, mutational model, distance-related detection, expectation of new genetic mutation, etc. In other words, no mutation model completely matches the behavior of microsatellites and no clear consensus over the accuracy of these measures to estimate genetic distances between populations, therefore both F_{st} and R_{st} measures, are often used together in microsatellite data analyses (27). Since both measures have their own drawbacks, we wanted these measures to complement each other, and so both methods were used in this study.

Genetic diversity parameters distinguished that DYS385a/b and DYS439 loci revealed the maximum and minimum number of alleles in both populations, respectively. Due to the lack of uniformity of the marker set used in different studies, the results are not very comparable. DYS385 was reported to show the most significant number of alleles in the Finnish population using 16 Y-STR loci (22). Also, the highest and lowest number of alleles were found for DYS458 and GATA-H4 markers, respectively, in ethnic Malaysian groups (Malay, Hindi, and Chinese) (28). Additionally, using a set of 11 Y-STR loci with demographic groups in America, the maximum number of alleles (18 alleles) was found in DYS385a/b markers, and the lowest number of alleles (5 alleles) was observed in DYS389I (29). Considering the DYS385a/b locus, the Shannon index values and the HE values were also calculated. The highest HE and Theta (θ) values were found for DYS385b and DYS385a loci in both populations. The comparative analysis, via F_{st} pairwise distance, revealed the statistical difference between the Fars and Isfahan populations. More than 8% of the 289 haplotypes were shared between and within the sub-populations. The American ethnic groups (Spanish, African, and European) were studied using 7 Y-STRs and determined that 85, 83, and 82% of the haplotypes were respectively observed only once in each population (30). The male population of Turkey was studied using 9 Y-STRs and most haplotypes were observed only once in the population (31). The average value of CE (0.6698), high value of HD, maximum (0.9725 in Fars), and minimum (0.8519 in Isfahan) DC values were compared with other corresponding studies. In another study, using 9 Y-STRs, the obtained HD and DC values were as 0.9938 and 8, respectively, in the Isfahan population, which is somewhat higher than the current results (32). Reports from other parts of Iran show that Tehran had HD and DC values of 0.997 and 0.900, respectively (32), while the HD and DC values were 0.9999 and 0.9884, respectively, in the Eastern region of Iran (33). Investigation of 17 Y-STRs in four Pakistan ethnic groups (Balochi, Pathans, Punjabis, and Sindhi) revealed the highest HD (HD=0.9957) for a Pathan population with 102 haplotypes (34). Using 27 Y-STRs, 102 haplotypes were identified in Kazakh populations and the HD and DC were 0.997 and 0.91, respectively (35). Furthermore, a total of 224 haplotypes were detected in two Mestizo populations, of which 98.66% were unique (36).

In addition, socio-geographical effects have been reported based on Y-STR diversity in different ethnic groups in America, including African, European, Spanish, Native, and Asian ethnic groups from various geographical regions. The highest DC was found in America's Asian population from two metropolitan areas, Tucson, AZ, and New York, NY. The Asian population in America also showed the most significant amount of HD (0.84) compared to the other ethnic groups (37). The average values of DC and HD were 0.9060 and 0.9994, respectively, in the European part of Russia (38). In a study on Indian population groups using 5 Y-SNPs markers, HD was in the range of 0.70 to 1.00 (39), while in another study, the average HD was 0.98 (40). To create a Y-STR database for the American population, the estimated DC values were 0.74, 0.75, and 0.69 for Hispanic, African, and native American populations, respectively (41). Using Y-SNPs and Y-STRs analysis, the HD and DC values were 0.9998 and 0.9680, respectively, in the Hui population of Liaoning province (42).

In terms of population genetic structure, our Y-STR data revealed minor differences between the two population groups, with the Isfahan population exhibiting lower levels of within-population diversity, higher levels of among-population differentiation, and a tendency toward a lower rate of demographic growth compared to the Fars population (Table S3). Comparing different Iranian populations with neighboring countries showed that the south Iranian populations have been influenced by eastern and north-western neighbors, with Pakistan having a larger effect on the south Iranian gene pool than Turkey (43). In a fascinating study, the northwest region of Iran was investigated using haplogroup G2-M406 (44). This region is renowned as the cradle of great civilizations like Mesopotamia and Elam, which is surrounded by the Zagros Mountains. It was calculated that the emergence of haplogroup G2-M406 in this region dates back more than 8,800 years. They highlighted that the gene flow from neighboring areas and the free movements of the local people had been sharply restricted by geography due to the Zagros Mountains. They also showed a westward migration of the Iranian population dwelling in this plateau to Europe through Turkey and the Caucasus (44). Haplogroup G has arisen due to a mutation from Haplogroup F, M201 in the case of Haplogroup G, and it has been reported that it has low frequencies in Western Europe (45). In haplogroup G, the segregation of G1 and G2 subdivisions occurred about 19,000 years ago, and the G2a and G2a3b subdivisions about 15,000 and 12,600 years ago, respectively. Although G2a was formed about 15,000 years ago, its contemporary subdivision expanded about 5,600 years ago from the Middle East into two branches: one to the Balkans and Western Europe, the other to the Caucasus and Pakistan (46). Various studies have addressed the origin of haplogroup G being in the regions between

eastern Anatolia, Armenia, and western Iran, with most of its current expansion being in the Middle East, the Caucasus, and southern Europe (Italy and Greece) (47).

Even having an average mutation rate of 0.2% per generation, Y-STRs can demonstrat the patrilineal affinities among subpopulations, and reveal the genetic composition and consanguinity scenarios (48), and are widely used to predict haplogroups with online tools. However, the occurrence of errors in Y-STR haplogroup prediction should not be overlooked. To this end, by increasing the number of Y-STRs employed to predict the haplogroup, correcting inaccurate prediction due to mutation is necessary using precise Y-SNP data. In other words, the existence of errors in Y-STR haplogroup predictions implies that validation using SNP analysis is appropriate when high accuracy is required in terms of haplogroup prediction (49). It has been shown that applying Y-SNPs in combination with Y-STR data can help shed light on the biogeographical origins of unknown male individuals with high precision, as this could mediate accessibility to robust phylogenetic trees and large-scale genome reference data (50). Therefore, more studies are needed to provide more insightful information about the geographic distribution of many Y-SNPs, and these sets of data could improve the geographic resolution of paternal ancestry inferences. One of the prime limitations of this study was the magnitude of subpopulation admixture and inter-ethnic structure. This matter wasn't considered in this study. Both inter-ethnic admixture and ancestral source subpopulations may contribute to fine scale Y-STRs heterogeneity within Iran ethnic groups. For the future genetic investigation of Iran's ethnic groups and subpopulation admixture, a panel of Y-SNPs should be considered, as it could infer distant evolutionary ancestry in comparison to Y-STRs that reflect close familial kinships.

Conclusion

This study covers comparative statistical analysis and various population genetics parameters from two socio-geographical populations of Iran (Fars and Isfahan) using 289 random unrelated samples and 17 Y-STRs as markers. The DYS385a/b marker revealed the maximum number of alleles in both studied populations. However, DYS439 and GATA-H4 markers in the Fars population and DYS19 and DYS439 markers in Isfahan individuals showed the minimum number of alleles. A high rate of haplotype diversity was observed in the total population and subpopulations of the Fars, and Isfahan regions. Haplotype discrimination capacity varied in Fars, Isfahan and the total populations. The results, again, showed the power of Y-STR markers for forensic national/regional human identification and haplotype comparison purposes.

Acknowledgements

We would like to express our gratitude to all the contributors, sample donors, and our colleagues in the forensic centers of the capital cities of Fars and Isfahan provinces. We are deeply indebted to Amin Mortazavi, Reza Raoofian, Reza Pazhoomand, Whit Athey, Farzan, Ghane Golmohammadi for their supports and help in this research.

Ethical Approval

The authors are committed to all ethical standards for human studies. This research protocol was under supervision by the Ethical Review Committees of Iranian Legal Medicine Research Center. The general formal ethical approval for this study was obtained from the Ethical Committee of the Legal Medicine Organization, Tehran (Approval ID: IR.LMO.REC.115021). Informed consent was obtained from all individual participants included in the study for using samples and publishing results.

References

1. Ellegren H. Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics. 2004;1;5(6):435-45.
2. Butler JM, Hill CR. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis Forensic Science Review. 2012;24(1):15.
3. McClintock JT. Using conventional STR technology in analyzing biological transfer evidence. InNext Generation Sequencing (NGS) Technology in DNA Analysis 2024 ;1 (pp. 21-34). Academic Press.
4. Vallian S, Moeini H. Genotyping of Five Polymorphic STR Loci in Iranian Province of Isfahan. Journal of Sciences, Islamic Republic of Iran. 2006;17(2):113-117
5. Agudo MM, Aanes H, Albert M, Janssen K, Gill P, Bleka Ø. An overview of autosomal STRs and identity SNPs in a Norwegian population using massively parallel sequencing. Forensic Science International: Genetics. 2024; 3:103057.
6. Soloveva EV, Skleimova MM, Minaycheva LI, Garaeva AF, Zhigalina DI, Churkin EO, Okkel YV, Timofeeva OS, Petrov IA, Seitova GN, Lebedev IN. PGT-M for spinocerebellar ataxia type 1: development of a STR panel and a report of two clinical cases. Journal of Assisted Reproduction and Genetics. 2024;41(5):1273-83.
7. Pedroza Matute S, Iyavoo S. Applications and Performance of Precision ID GlobalFiler NGS STR, Identity, and Ancestry Panels in Forensic Genetics. Genes. 2024;28;15(9):1133.
8. Costa R, Fadoni J, Amorim A, Cainé L. Y-STR Databases—Application in Sexual Crimes. Genes. 2025;

25:16(5):484.
9. Jobling MA, Pandya AR, Tyler-Smith CH. The Y chromosome in forensic analysis and paternity testing. International Journal of Legal Medicine. 1997; 110:118-24.
10. Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. Nature Reviews Genetics. 2003;1;4(8):598-612.
11. Ahmady K. Ethnic, national and identity demand in Iran on the axis of justice and development: A grounded theory method study among five major ethnic groups. European Journal of Humanities and Social Sciences. 2023;31;3(5):41-54.
12. Roewer L, Willuweit S, Stoneking M, Nasidze I. A Y-STR database of Iranian and Azerbaijanian minority populations. Forensic Science International: Genetics. 2009; 1;4(1):e53-5.
13. Sayyari M, Salehzadeh A, Tabatabaiefar MA, Abbasi A. Profiling of 17 Y-STR loci in Mazandaran and Gilan provinces of Iran. Turkish Journal of Medical Sciences. 2019;49(5):1277-86.
14. Sayyari M, Salehzadeh A, Tabatabaiefar MA, Ali A. Genetic polymorphisms of Y-chromosome short tandem repeats (Y-STRs) in a male population from Golestan province, Iran. Molecular Biology Research Communications. 2020;9(1):11.
15. Mehrjoo Z, Fattahi Z, Beheshtian M, Mohseni M, Poustchi H, Ardalani F, et al. Distinct genetic variation and heterogeneity of the Iranian population. PLoS Genetics. 2019;24;15(9): e1008385.
16. Willuweit S, Roewer L. The new Y chromosome haplotype reference database. Forensic Science International: Genetics. 2015;1; 15:43-8.
17. Nei M. Molecular evolutionary genetics. Columbia university press; 1987;31.
18. Peakall RO, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 2006;6(1):288-95.
19. Moges AD, Admassu B, Belew D, Yesuf M, Njuguna J, Kyalo M, et al. Development of microsatellite markers and analysis of genetic diversity and population structure of Colletotrichum gloeosporioides from Ethiopia. PloS One. 2016;15;11(3): e0151257.
20. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;1;131(2):479-91.
21. Robino C, Crobu FR, Di Gaetano C, Bekada A, Benhamamouch S, Cerutti N, et al. Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. International Journal of Legal Medicine. 2008; 122:251-5.
22. Hedman M, Pimenoff V, Lukka M, Sistonen P, Sajantila A. Analysis of 16 Y STR loci in the Finnish population reveals a local reduction in the diversity of male lineages. Forensic Science International. 2004;28;142(1):37-43.
23. Alves C, Gusmão L, Meirinhos J, Amorim A. Making the most of Y-STR haplotypes: The HapYDive. In International Congress Series 2006 Apr 1 (Vol. 1288, pp. 201-203). Elsevier.
24. Athey TW. Haplogroup prediction from Y-STR values using an allele-frequency approach. Journal of Genetics and Genomics. 2005; 1:1-7.
25. Athey TW. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. Journal of Genetics and Genomics. 2006;2(2):34-9.
26. Oh A, Oh BU. Genetic differentiation that is exceptionally high and unexpectedly sensitive to geographic distance in the absence of gene flow: Insights from the genus Eranthis in East Asian regions. Ecology and Evolution. 2022;12(6): e9007.
27. Alinaghi S, Mohseni M, Fattahi Z, Beheshtian M, Ghodratpour F, Ashrafi FZ, Arzhangi S, Jalalvand K, Najafipour R, Khorshid HR, Kahrizi K. Genetic Analysis of 27 Y-STR Haplotypes in 11 Iranian Ethnic Groups. Archives of Iranian Medicine. 2024 Feb;27(2):79.
28. Chang YM, Perumal R, Keat PY, Kuehn DL. Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. Forensic Science International. 2007;22;167(1):70-6.
29. Shewale JG, Bhushan A, Nasir H, Schneida E, Washington B, Fleming A, et al. Population Data for Four Population Groups from the United States for the Eleven Y-Chromosome STR Loci Recommended by SWGDAM. Journal of Forensic Sciences. 2006;1;51(3).
30. Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, et al. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. The American Journal of Human Genetics. 2001;1;68(4):990-1018.
31. Henke J, Henke L, Chatthopadhyay P, Kayser M, Dulmer M, Cleef S, et al. Application of Y-chromosomal STR haplotypes to forensic genetics. Croatian Medical Journal. 2001;1;42(3):292-7.
32. Ikram MS, Mehmood T, Rakha A, Akhtar S, Khan MI, Al-Qahtani WS, Safhi FA, Hadi S, Wang CC, Adnan A. Genetic diversity and forensic application of Y-filer STRs in four major ethnic groups of Pakistan. BMC Genomics. 2022;30;23(1):788.
33. Nasidze I, Schädlich H, Stoneking M. Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. Forensic science international. 2003;14;137(1):85-93.
34. Tabrizi AA, Hedjazi A, Kerachian MA, Honarvar Z, Dadgarmoghaddam M, Raoofian R. Genetic profile of 17 Y-chromosome STR haplotypes in East of Iran. Forensic Science International: Genetics. 2015;1;14: e6-7.
35. Ashirbekov Y, Zhunussova A, Abaildayev A, Bukayeva A, Sabitov Z, Zhabagin M. Genetic polymorphism of 27 Y-STR loci in Kazakh populations from Central Kazakhstan. Annals of Human Biology. 2024; 31:51(1):2377571.
36. Sánchez-Méndez AD, Narvaez-Rivera SE, Rangel-Villalobos H, Hernández-Bello J, López-Quintero A, Moreno-Ortíz JM, Ramos-González B, Aguilar-Velázquez JA. Genetic Diversity and Forensic Parameters of 27 Y-STRs in Two Mestizo Populations from Western Mexico. Genes. 2025; 19:16(3):352.
37. Hammer MF, Chamberlain VF, Kearney VF, Stover D, Zhang G, Karafet T, et al. Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. Forensic Science International. 2006;1;164(1):45-55.
38. Roewer L, Willuweit S, Krüger C, Nagy M, Rychkov S,

Morozowa I, et al. Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. International Journal of Legal Medicine. 2008; 122:219-23.
39. Cordaux R, Bentley G, Aunger R, Sirajuddin SM, Stoneking M. Y-STR haplotypes from eight south Indian groups based on five loci. Journal of Forensic Sciences. 2004;1;49(4):847-8.
40. Ayadi I, Ammar-Keskes L, Rebai A. Haplotypes for 13 Y-chromosomal STR loci in South Tunisian population (Sfax region). Forensic Science International. 2006;20;164(2-3):249-53.
41. Kayser M, Brauer S, Schädlich H, Prinz M, Batzer MA, Zimmerman PA, Boatin BA, Stoneking M. Y chromosome STR haplotypes and the genetic structure of US populations of African, European, and Hispanic ancestry. Genome Research. 2003;1;13(4):624-34.
42. Fu D, Adnan A, Yao J, Aldayan NH, Wang CC, Hongyi C. Unraveling the paternal genetic structure and forensic traits of the Hui population in Liaoning Province, China using Y-chromosome analysis. BMC Genomics. 2023;17;24(1):691.
43. Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ. Iran: tricontinental nexus for Y-chromosome driven migration. Human Heredity. 2006;16;61(3):132-43.
44. Bahmanimehr A, Nikmanesh F, Yepiskoposyan L. Paternal genetic landscape of Zagros region of Iran and its role in the gene flow in the populations of the region. Iranian Journal of Epidemiology. 2014;10(3):43-53.
45. Goff PG, Athey TW. Diagnostic Y-STR markers in haplogroup G. Journal of Genetic Genealogy. 2006;2(1):12-7.
46. Rootsi S, Myres NM, Lin AA, Järve M, King RJ, Kutuev I, et al. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. European Journal of Human Genetics. 2012;20(12):1275-82.
47. Balanovsky O, Zhabagin M, Agdzhoyan A, Chukhryaeva M, Zaporozhchenko V, Utevska O, et al. Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y-chromosome and reveals migrations of Iranic speakers. PLoS One. 2015;7;10(4):e0122968.
48. Tariq M, Ahmad H, Hemphill BE, Farooq U, Schurr TG. Contrasting maternal and paternal genetic histories among five ethnic groups from Khyber Pakhtunkhwa, Pakistan. Scientific Reports. 2022;19;12(1):1027.
49. Petrejčíková E, Hronská D, Bernasovská J, Boroňová I, Gabriková D, Bôžiková A, et al. Y-SNP analysis versus Y-haplogroup predictor in the Slovak population. Anthropologischer Anzeiger. 2014;1;71(3).
50. Roewer L. Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. Wiley Interdisciplinary Reviews: Forensic Science. 2019;1(4): e1336.

Table S1. Each haplotype information used in current study

Region	Number of haplotypes	Accession number	Source
Fars[1]	109	YA004229	YHRD[2] – Current study
Isfahan[3]	180	YA004228	YHRD – Current study
Center of Iran	154	YA003782	YHRD
Mashhad[4]	127	YA003903	YHRD
Zahedan[5]	102	YA003901	YHRD
Pakistan[6]	269	YA003846	YHRD
Turkey[7]	296	YA002964	YHRD
Iraq[8]	124	YA003858	YHRD

[1] Fars province, one of the southern provinces of Iran, had 109 Y-DNA samples in current study. Fars province, the capital of ancient Iran, with 122608 Km^2 total area, and a population of 4851274 individuals (national statistic center of Iran in 2016 census, https://www.amar.org.ir/english) is the 4th most populated province.
[2] Y-Chromosome Haplotype Reference Database (YHRD - https://yhrd.org/)
[3] Isfahan province, one of the central provinces of Iran, had 180 Y-DNA samples in this study. Isfahan province with 107029 Km^2 total area, and a population of 5120850 individuals (national statistic center of Iran in 2016 census, https://www.amar.org.ir/english) is the 3rd populated province.
[4] Mashhad city is the capital city of Khorasan-e-Razavi province, one of the eastern provinces of Iran. Khorasan-e-Razavi province with 118884 Km^2 total area, and a population of 6434501 individuals (national statistic center of Iran in 2016 census, https://www.amar.org.ir/english) is the 2nd populated province.
[5] Zadehan city is the capital city of Sistan and Baluchestan province, one of the eastern provinces of Iran. Sistan and Baluchestan province with 181785 Km^2 total area, and a population of 27750414 individuals (national statistic center of Iran in 2016 census, https://www.amar.org.ir/english) is the 10th populated province.
[6] Pathan, Pakistan. Pakistan is one of the eastern neighbors of Iran with 959 km border line.
[7] Turkey is the north-western neighbor of Iran with 534 km border line.
[8] Iraq is the western neighbor of Iran with 1599 km border line.

Table S2. Haploid allelic frequencies of 17 STR in Fars and Isfahan.

Locus	Allele	Fars	Isfahan
DYS19	13	0.138	0.078
	14	0.404	0.472
	15	0.303	0.317
	16	0.128	0.111
	17	0.028	0.022
DYS385a	9	0.018	0.033
	10	0.028	0.017
	11	0.174	0.189
	12	0.165	0.150
	13	0.220	0.317
	14	0.165	0.150
	15	0.110	0.067
	16	0.046	0.044
	17	0.064	0.022
	18	0.009	0.011
DYS385b	11	0.018	0.011
	12	0.018	0.028
	13	0.037	0.022
	14	0.183	0.178
	15	0.156	0.167
	16	0.174	0.211
	17	0.202	0.167
	18	0.073	0.133
	19	0.083	0.039
	20	0.028	0.011
	21	0.028	0.033
DYS389I	9	0.009	0.000
	10	0.009	0.000
	12	0.165	0.150
	13	0.596	0.65

Table S2. Haploid allelic frequencies of 17 STR in Fars and Isfahan.

Locus	Allele	Fars	Isfahan
DYS389II	25	0.009	0.000
	27	0.009	0.000
	28	0.083	0.072
	29	0.294	0.417
	30	0.376	0.361
	31	0.202	0.128
	32	0.028	0.022
DYS390	20	0.000	0.006
	21	0.000	0.033
	22	0.220	0.150
	23	0.358	0.400
	24	0.248	0.278
	25	0.165	0.128
	26	0.009	0.006
DYS391	9	0.055	0.056
	10	0.743	0.706
	11	0.147	0.233
	12	0.009	0.006
	13	0.028	0.000
	14	0.018	0.000
DYS392	9	0.018	0.000
	10	0.046	0.044
	11	0.688	0.600
	12	0.046	0.050
	13	0.083	0.150
	14	0.083	0.122
	15	0.028	0.033
	16	0.009	0.000
DYS393	10	0.000	0.006
	11	0.018	0.039
	12	0.385	0.467
	13	0.459	0.389
	14	0.083	0.078
	15	0.055	0.022
DYS437	14	0.560	0.406
	15	0.266	0.450
	16	0.147	0.144
	17	0.018	0.000
	19	0.009	0.000
DYS438	9	0.312	0.378
	10	0.440	0.267
	11	0.211	0.228
	12	0.037	0.128
DYS439	10	0.193	0.139
	11	0.413	0.389
	12	0.358	0.300
	13	0.037	0.150
	14	0.000	0.022
DYS448	14	0.009	0.000
	16	0.000	0.006
	18	0.046	0.050
	19	0.275	0.378
	20	0.450	0.306
	21	0.165	0.228
	22	0.046	0.033
	23	0.009	0.000
DYS456	13	0.009	0.033
	14	0.138	0.094
	15	0.569	0.539
	16	0.257	0.217
	17	0.018	0.100
	18	0.000	0.017
	22	0.009	0.000

Table S2. Haploid allelic frequencies of 17 STR in Fars and Isfahan.

Locus	Allele	Fars	Isfahan
DYS458	14	0.055	0.028
	15	0.193	0.267
	16	0.294	0.389
	17	0.202	0.172
	18	0.138	0.094
	19	0.083	0.044
	20	0.028	0.006
	22	0.009	0.000
DYS635	19	0.000	0.006
	20	0.083	0.056
	21	0.459	0.311
	22	0.147	0.161
	23	0.193	0.333
	24	0.083	0.094
	25	0.037	0.033
	32	0.000	0.006
GATA H4	10	0.064	0.050
	11	0.468	0.456
	12	0.385	0.350
	13	0.083	0.144

Table S3. Number of alleles, gene diversity and theta in Fars and Isfahan populations, Iran.

Locus	Number of alleles[1]		Shanon index		Expected heterozygosity		Theta (θ)	
	Fars	Isfahan	Fars	Isfahan	Fars	Isfahan	Fars	Isfahan
DYS19	5(3.439)	3(2.062)	1.363	1.246	0.71577	0.66151	5.68897	3.86405
DYS385a	10(6.539)	11(6.509)	2.008	1.883	0.85491	0.81515	23.25170	14.13227
DYS385b	11(6.904)	4(3.546)	2.092	2.025	0.86306	0.85109	26.16493	22.04761
DYS389I	6(2.338)	5(3.067)	1.063	0.886	0.57764	0.51788	2.30290	1.65107
DYS389II	7(3.621)	7(3.608)	1.441	1.270	0.73055	0.67772	6.38659	4.31384
DYS390	5(3.771)	4(1.801)	1.387	1.441	0.74159	0.72688	6.98776	6.20279
DYS391	6(1.730)	6(2.481)	0.877	0.775	0.42593	0.44711	1.01717	1.13568
DYS392	8(2.031)	6(2.652)	1.167	1.250	0.51223	0.60031	1.60157	2.62986
DYS393	5(2.709)	10(5.280)	1.164	1.161	0.63677	0.62638	3.28962	3.08189
DYS437	5(2.463)	6(3.429)	1.075	1.005	0.59956	0.61558	2.61811	2.88344
DYS438	4(2.967)	5(3.527)	1.174	1.320	0.66905	0.72197	4.06491	5.96843
DYS439	4(2.968)	3(2.578)	1.172	1.372	0.66922	0.72048	4.06960	5.89966
DYS448	7(3.231)	6(2.796)	1.381	1.359	0.69691	0.71235	4.94278	5.54295
DYS456	6(2.445)	7(3.792)	1.102	1.299	0.59650	0.64593	2.57102	3.48843
DYS458	8(5.163)	8(4.048)	1.780	1.512	0.81380	0.74041	13.92080	6.91981
DYS635	6(3.520)	4(2.830)	1.490	1.578	0.72256	0.75717	5.99588	7.97937
GATA H4	4(2.643)	3(2.062)	1.105	1.155	0.62742	0.65022	3.10191	3.58670
Mean	6.294(3.440)	5.882(3.340)	1.344	1.326	0.67300	0.67500	1.77400	1.78200

[1] Values in parentheses present effective numbers of alleles.

Table S4. R_{st} values obtained for comparison between five Iranian and three regional populations. See Table S1 for complementary information about each population. The off-diagonal entries in the matrix provide the R_{st} and *P*-value between each variable pair. The values above and below the diagonal are *P*-values and R_{st} values, respectively.

Population	Fars	Isfahan	Central Iran	Mashhad	Zahedan	Pakistan	Turkey	Iraq
Fars	-	0.0449	0.0126	0.0247	0.1479	0.0000	0.0179	0.0722
Isfahan	0.0079	-	0.7718	0.3732	0.1745	0.0000	0.0492	0.0631
Central Iran	0.0130	-0.0024	-	0.3669	0.1372	0.0000	0.0034	0.0447
Mashhad	0.0117	0.0002	0.0004	-	0.2293	0.0000	0.0410	0.1590
Zahedan	0.0046	0.0031	0.0044	0.0027	-	0.0000	0.0131	0.0720
Pakistan	0.0856	0.0773	0.0783	0.0462	0.0692	-	0.0000	0.0000
Turkey	0.0085	0.0044	0.0104	0.0057	0.0108	0.0581	-	0.5088
Iraq	0.0085	0.0071	0.0093	0.0041	0.0087	0.0518	0.0008	-

Table S5. The F_{st} values obtained for comparison between five Iranian and three regional populations (See Table S1 for complementary information about each population). The off-diagonal entries in the matrix provide the F_{st} and *P*-value between each variable pair. The values above and below the diagonal are *P*-values and F_{st} values, respectively.

Population	Fars	Isfahan	Central Iran	Mashhad	Zahedan	Pakistan	Turkey	Iraq
Fars[1]	-	0.0017	0.0010	0.0812	0.1256	0.0001	0.0003	0.0616
Isfahan[2]	0.0008	-	1.0000	0.0079	0.0513	0.0000	0.0000	0.0050
Central Iran	0.0011	-0.0033	-	0.0096	0.0244	0.0000	0.0000	0.0032
Mashhad[3]	0.0002	0.0006	0.0007	-	1.0000	0.0000	1.0000	0.1893
Zahedan[4]	0.0003	0.0005	0.0007	0.0000	-	0.0000	1.0000	0.4948
Pakistan[5]	0.0052	0.0056	0.0058	0.0048	0.0049	-	0.0000	0.0002
Turkey[6]	0.0003	0.0007	0.0009	0.0000	0.0000	0.0049	-	0.0606
Iraq[7]	0.0004	0.0007	0.0011	0.0001	0.0000	0.0050	0.0001	-

Journal of Sciences, Islamic Republic of Iran 36(1): 21 - 27 (2025) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Antiretroviral Therapy Among HIV-Infected Pregnant Women on Their Offspring

M. Montazeri[1], M.A. Davarpanah[2], M. Montazeri[3], L. Davoodi[4*]

[1] *Razi Clinical Researches Development, Mazandaran University of Medical Sciences, Sari, Islamic Republic of Iran*
[2] *Department of Infectious and Tropical Diseases, Faculty of Medicine, Shiraz University of Medical Sciences, Shiraz, Islamic Republic of Iran*
[3] *Sarem Fertility and Infertility Research Center (SAFIR), Sarem Women's Hospital, Tehran, Islamic Republic of Iran*
[4] *Department of Infectious Diseases, Antimicrobial Resistance Research Center, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Islamic Republic of Iran*

Received: 2 September 2023 / Revised: 30 October 2025 / Accepted: 31 December 2025

Abstract

The risk of mother-to-child transmission (MTCT) of human immunodeficiency virus (HIV) infection is approximately 30%. However, antiretroviral drugs can reduce MTCT to less than 2%. This study was designed to determine the effect of antiretroviral therapy among HIV-infected women and its maternal and neonatal outcomes in Iran. The study is a retrospective analysis of mother-infant data from Shiraz, Southern Iran, between 2006 and 2012. HIV-infected pregnant women were divided into two groups of intervention (receiving treatment or chemoprophylaxis) and control (not receiving any treatment). Maternal and neonatal information were extracted and recorded. The data were entered into SPSS software and were analyzed. The mother-to-child transmission was 2.9% in the intervention group compared to 15.8% in the control group (OR=0.01, 95% CI: 0.002-0.125, p<0.0001). The infant HIV infection rate was significantly higher in male infants (OR=2.76, CI 95% 2.213-3.327), NVD delivery (OR=3.78, CI 95% 3.140-4.409), and breastfeeding (OR=26, CI 95% 7.87-85.90). Treatment intervention significantly reduces the HIV transmission from infected mothers to their infants. However, the rate of vertical transmission in Iran remains higher than those reported in developed countries despite treatment interventions, and additional preventive measures appear necessary.

Keywords: Chemoprophylaxis; Treatment; Pregnancy; HIV; Mother-to-Child Transmission.

Introduction

About half of the 33.4 million people infected with HIV (Human Immunodeficiency Virus) in the world are women, and about half a million children infected with HIV are infected through vertical transmission (1, 2). If a woman infected with HIV becomes pregnant, there is a chance of 35% of mother-to-child transmission (MTCT). Perinatal HIV infection can occur during pregnancy, at delivery, or during lactation. About 9% of the infants who were HIV-negative at birth would be infected after 18 months through breastfeeding (3, 4). It is estimated

* Corresponding Author: Tel: +989113270253 ; Email: lotfdavoodi@yahoo.com

Perinatal HIV infection can occur during pregnancy, at delivery, or during lactation. About 9% of the infants who were HIV-negative at birth would be infected after 18 months through breastfeeding (3, 4). It is estimated that 200,000 infants annually in the world become infected with HIV through mother-to-child transmission (5).

Strategies for the treatment of pregnant women with HIV in the pharmacological area and recognition of the ways to prevent perinatal and during labor transmission have made considerable progress. In the US and Europe, using antiviral drugs has dramatically decreased the risk of mother-to-child transmission of HIV infection. Among factors with a key role in such a reduction, universal screening of pregnant women with HIV, cesarean delivery, and if possible, avoiding breastfeeding by mothers can be mentioned (6, 7).

However, with antiretroviral drugs and effective viral suppression, the risk of infants' infection through mother-to-child transmission has declined to less than 2% in Europe and the United States (2, 4). It is thus recommended that all pregnant women with HIV, regardless of CD4 count and HIV RNA level, receive antiviral prophylaxis to prevent mother-to-child transmission. The post-exposure prophylaxis (PEP) in infants has also been suggested to reduce the risk of HIV acquisition. Post-exposure prophylaxis can protect the infant against the virus that may have been transmitted to the embryo during labor or during passage through the birth canal (8-11). Antiviral prophylaxis in infants whose mothers have received HIV care late in pregnancy is very important. These women, even in cases of receiving the treatments in the best conditions, may be detected at high risk of viremia, which may last for weeks or months to be suppressed (12-14).

Some data suggest that a combination of antiviral drugs with the aim of complete viral suppression can decrease the risk of vertical transmission of HIV infection. The data also showed that a woman with no detectable viral count, despite the low risk, can still transmit the HIV infection to her child. Thus, regardless of the number of viruses, antiviral prophylaxis is recommended for all pregnant women to prevent mother-to-child transmission (15).

Data available on HIV infection in Iran are different. Sedaghat et al. (16) estimated that the number of infected patients will increase from 88,000 in 2010 to 105,000 in 2014. According to the latest report of HIV cases recorded by the end of 2013 in Iran, which was provided according to statistics collected from the Universities of Medical Sciences and Health Services, a total of 26125 people with HIV were identified in the country. The mother-to-child transmission accounts for 3.2% of cases (17).

No study has been done on the vertical transmission of HIV and the effect of prophylaxis on infants' infection in Iran so far. Also, similar studies have been conducted in limited countries in recent years. Therefore, this study was designed to determine the effect of chemoprophylaxis or therapy in women with HIV and its maternal and neonatal outcomes.

Materials and Methods

The study is a retrospective analysis of mother-infant data from the records of the Shiraz Counseling Center for Behavioral Disorders. Eligible subjects were HIV-infected pregnant women who accessed the service between April 2006 and March 2012 and delivered newborns. Their infants were followed up for 18 months. Other subjects who HIV diagnosed in their infants in follow-up period were also included in the study. Infants with persistence of antibody beyond 18 months of age and/or a positive virological marker of infection, regardless of age, were considered HIV-infected. The research protocol was previously approved by the Ethical Committee of Shiraz University of Medical Sciences (Project number 3850/2012).

According to the protocol of Iran's AIDS Society, pregnant women with HIV should receive treatment or chemoprophylaxis; however, in some circumstances, such as lack of prenatal detection or lack of mothers' cooperation, this may not happen. Mothers who used the IMOD or lacked cooperation in referring to the center and in completing the questionnaire were excluded from the study.

Accordingly, the study population was divided into two groups: HIV-infected pregnant women who received treatment or prophylaxis, defined as the intervention group, and HIV-infected pregnant women who received no treatment or chemoprophylaxis, or received incomplete treatment or chemoprophylaxis, defined as the control group. Mothers whose diagnosis of HIV was made after their infant's infection was confirmed were also included in the control group.

Information collected includes: demographic and epidemiologic characteristics of the pregnant women, age at pregnancy, smoking, use of Methadone, hookah, opium, heroin and alcohol, other diseases (such as diabetes, hepatitis C, Hepatitis B, tuberculosis, hypertension), hemoglobin concentration, gravidity and parity, educational level, place of residence (urban or rural), type of delivery (normal vaginal delivery or cesarean), gestational age at delivery, number of CD4, time of diagnosis (before pregnancy, during pregnancy, perinatal and postnatal), clinical stage

according to the WHO, in case of receiving intervention, the type and duration of intervention, the reason for not-receiving the treatment or chemoprophylaxis. Also, the information on newborns, such as the following, was recorded: Sex, birth weight, prematurity, infant feeding (breast milk or formula), survival status (dead or alive newborn), diagnosis time of neonatal infection with HIV, history of hospitalization from birth to 18 months, and the follow-up period after birth.

Data were cleaned, edited, entered into a computer, and analyzed using the Statistical Package for the Social Sciences (SPSS) version 18.0 (SPSS Inc., Chicago, IL). The data obtained were described using frequency tables and corresponding graphs. To describe qualitative properties, frequencies and percentages were used, while mean and standard deviation were used for quantitative traits. To determine the relationship between qualitative variables, the chi-square test and the Fisher exact test were used, and quantitative variables were analyzed using the t-test. The T-Test was used to confirm normal distribution, while the Mann-Whitney test was used when normality was rejected. The statistical significance level was set at 0.05 for all tests, so that P-values < 0.05 indicate statistically significant differences.

Results

HIV-infected Pregnant Women

Of the 81 pregnant women who participated in this study, 34 and 47 were assigned to the intervention and control groups, respectively. The mean age of the women was 29.21±5.77 years in the intervention group and 28.27±4.57 years in the control group, with no significant difference (P=0.422). There was no significant difference in gravidity or parity between the intervention and control groups (P=0.158 and P=0.845, respectively). Regarding educational status, in the intervention group, 2 patients were illiterate, 20 patients had an education level below diploma, and 14 patients had a college or higher level of education. Whereas, in the controls, there were 4 illiterate people, 30 people with an education level below diploma, and 13 subjects with a college or higher level of education. There was no significant difference between the two groups regarding educational level (P = 0.123). In the intervention group, 24 women were urban residents, and 10 were rural residents. However, in the control group, 31 women lived in urban areas, whereas 16 lived in rural areas. Two study groups had no significant difference regarding the place of residence (urban or rural) (P=0.660). Regarding prenatal care, the women in the two groups did not differ, with 37 patients (78.7%) in the control group and 32 patients (94.1%) in the intervention group receiving prenatal care (P = 0.065).

The mean hemoglobin level before delivery was 11.61±1.44 mg/dl and 11.63±0.96 mg/dl in the intervention and control groups, respectively, and no significant difference was observed between the two groups (P=0.932). The mean CD4 cell count before delivery was 373.44±173.45 cells/mm3 in the intervention group and 352.20±148.23 cells/mm3 in the control group, with no significant difference (P=0.682).

Cesarean delivery was significantly higher in the intervention group, so that 70.6% of the women in this group had cesarean section, while the frequency of this type of delivery in the control group was 46.8% (P=0.033). The mean gestational age in the intervention group was 37.06±2.41 weeks, while it was 38.48±1.59 weeks in the control group, a difference was significantly lower in the intervention group (P=0.002).

Diagnosis of HIV infection in the control group was as follows: 6 subjects (12.8%) before pregnancy, 7 patients (14.9%) during pregnancy, 6 patients (12.8%) during labor and 28 cases (59.5%) after delivery. Diagnosis of infection in the intervention group was as follows: 16 subjects (47.1%) before pregnancy and 18 patients (52.9%) during pregnancy (P<0.0001). Only 3 patients had received HIV treatment before pregnancy and all of them were in the intervention group.

According to the WHO criteria (in patients diagnosed before delivery), in the control group, 10 patients (76.9%) were in stage I, and 3 patients (27.1%) were in stage III. No patient in this group was in stages II and IV. In contrast, in the intervention group, 31 patients (91.3%) were in stage I, 1 (2.9%) in stage II, 1 (2.9%) in stage III, and 1 (2.9%) in stage IV. The difference in disease stage between the two groups was not significant according to WHO criteria (P=0.152). In the control group, 34 patients (72.3%) were not treated due to lack of knowledge about infection, and 13 patients (27.7%) were not treated due to lack of cooperation. In the intervention group, 22 patients (64.7%) had received prophylaxis, and 12 patients (35.3%) had been treated.

Characteristics of the Newborns

In the control group, 28 newborns were male and 19 were female, whereas in the intervention group, 13 were male and 21 were female. The two groups did not differ regarding the gender of newborns (Table 1). The mean birth weight of newborns in the intervention and control groups was 2794.12±579.28 g and 3005.53±475.82 g, respectively; however, the t-test showed no significant difference in weight between infants born to women in both groups (P=0.076). Prematurity was observed in 26.5% of infants in the intervention group and 14.9% in

Table 1. Comparison of characteristics and outcomes of the newborns of HIV-infected pregnant women between the two intervention and control groups

		Intervention group N (%)	Control group N (%)	P value
Gender	Male	13 (38.2)	28 (59.6)	0.058*
	Female	21 (61.8)	19 (40.4)	
Prematurity	+	9 (26.5)	7 (14.9)	0.260**
	-	25 (73.5)	40 (85.1)	
Infant survival	Alive	33 (97.1)	41 (87.2)	ns†
	Dead	1 (2.9)	6 (12.8)	
Breastfeeding	+	0 (0)	33 (70.2)	<0.0001**
	-	34 (100)	14 (28.8)	
Progression to HIV	+	1 (2.9)	3 (15.8)	<0.0001**
	-	33 (97.1)	16 (84.2)	
Hospitalization	+	3 (8.8)	22 (46.8)	0.031**
	-	31 (91.2)	25 (53.2)	

*chi-square, **Fisher exact, †ns=not significant

Table 2. Neonatal and maternal factors associated with newborns' HIV infection

		HIV infected newborn		P value*
		+	-	
Time of diagnosis	Before pregnancy	2 (6.3)	20 (40.8)	<0.0001
	During prenatal care	1 (3.1)	24 (49)	
	During delivery	1 (3.1)	5 (10.2)	
	Following delivery	28 (87.5)	0 (0)	
Gender	Male	21 (65.6)	20 (40.8)	0.029
	Female	11 (34.4)	29 (59.2)	
Type of delivery	NVD	20 (62.5)	15 (30.6)	0.005
	C/S	12 (37.5)	34 (69.4)	
breastfeeding	+	26 (81.2)	7 (14.3)	<0.0001
	-	6 (18.8)	42 (85.7)	

*chi-square test

the control group, with no significant difference (Table 1). One case of a dead baby in the intervention group (2.9%) and 6 cases in the control group (12.8%) were found (Table 1). The cause of death in all 6 cases of the control group was reported as HIV, but in the intervention group, the only cause was due to congenital heart disease (CHD). None of the infants in the intervention group were breastfed, whereas in the non-intervention group, 28.8% of infants were breastfed (Table 1).

At the 18-month follow-up, the admissions cases were reported as 3 in the intervention group and 22 in the control group (Table 1). The causes of being hospitalized in the control group were as follows: seven infants due to respiratory infections, 3 with diarrhea and vomiting, 3 due to low platelet count, 2 with jaundice and hepatomegaly, 2 due to Failure to thrive (FTT) and lung infections, 2 infants of FUO (fever with unknown origin), an infant with oral thrush, infectious diarrhea and BCG lymphadenopathy, one case of jaundice, and one with Bacille Calmette-Guérin (BCG) lymphadenitis. However, the cause of admission in the intervention group included two cases of pulmonary infection and one CHD case.

Mother-to-child Transmission of HIV

The mother-to-child transmission of HIV infection was significantly higher in the control group (2.9% in the intervention group versus 15.8% in the control group, P<0.0001) (Table 1). The odds ratio (OR) of MTCT of HIV was 0.01 (95% CI: 0.002-0.125) among mothers who received any intervention compared to mothers who received no treatment or chemoprophylaxis.

According to Table 2, the neonatal HIV infection rate in mothers with prenatal diagnosis was lower (P<0.0001). Also, the infant HIV infection rate was significantly higher in male infants (OR=2.76, CI 95% 2.213-3.327), NVD delivery (OR=3.78, CI 95% 3.140-4.409), and breastfeeding (OR=26, CI 95% 7.87-85.90).

Discussion

Based on the results of this study, the rate of mother-to-child transmission in women with intervention was 2.9%, and in women who received no treatment intervention, it was 15.8%. According to the latest report

of recorded HIV cases in Iran in 2013, based on statistics collected from the Universities of Medical Sciences and Health Services, the mother-to-child transmission rate accounted for 3.2% of all HIV infection cases (17). However, there are no reports of the MTCT rate in the population of pregnant women in Iran, and this is the first study about mother-to-child transmission of HIV in this country. Therefore, it is not possible to compare these results with other data presented in Iran.

In comparing the results of this study with studies in other countries, in Lussiana et al. (18) the study in Angola, 68 women (65.4%) received antiviral combination therapy, but the other 36 women (34.6%) had not been treated due to being diagnosed after labor. The transmission rate from mother to child in the intervention and non-intervention groups was reported as 8.5% and 38.9%, respectively. In a cohort study in South Africa by Chetty et al. (19) among 260 infants born to HIV-pregnant women who were under Prevention of Mother-to-Child Transmission (PMTCT) programs, the rate of transmission from mother to child was 2.7% (6 cases). In Goswami and Chakravorty study (20), which was conducted by 4 years survey in India, despite enrolling 248 women who received treatment for HIV infection in the study, only 46 newborn babies were assessed due to the inability to follow up. Of these 46 cases, mother-to-child transmission was reported in only one case; according to the exclusion of many subjects from the study, the results are not comparable.

In the most recent study performed over a 6-year period in Ethiopia by Koye and Zeleke (21), among a total of 509 infants from mothers receiving antiretroviral treatment, there were 51 cases (10%) of MTCT. Neubert et al. (22) evaluated 118 mother-infant pairs treated for HIV therapy in pregnancy and in HIV exposed newborns from 2000 to 2010. Based on the results of this study, the overall Transmission risk in the group, regardless of risk factors and completion of prophylaxis, was 1.8%. If transmission prophylaxis was complete, transmission risk was 0.9%. In two studies in Brazil, the rate of vertical transmission in all female patients (treated and untreated) was much lower than in our study (9.9% and 9.7%, respectively) (23, 24). In Warszawski et al.'s study (25) in France, in women with HIV infection who had received an antiretroviral regimen during pregnancy, the vertical transmission rate was 1.3%. Also, another study in the United Kingdom and Ireland by Townsend et al. (26), the overall rate of mother-to-child transmission was 1.2%. In both studies, the rate was lower than the transmission rate in our study.

Data suggest that the combination of antiviral drugs with the goal of complete suppression of the virus can reduce the risk of perinatal transmission of HIV infection (15). Pharmacologic intervention with antiviral drugs during pregnancy follows two separate but related purposes: Treatment to reduce perinatal transmission of HIV and treatment of the mother infection if needed. Women with no need for treatment are also recommended to receive antiviral prophylaxis to prevent the mother-to-child transmission of HIV infection (8). With prophylaxis and effective viral suppression, the risk of infant infection through MTCH has declined to less than 2% in Europe and the United States (1). In our study, the transmission rate was higher than 2%; however, this can be attributed to the smaller number of patients undergoing intervention. In our study, only one infant of a woman undergoing medical treatment developed HIV infection, but given the small sample size, the vertical transmission rate was calculated to be more than 2 %. Thus, the vertical transmission rate of 2.9% can be attributed to the smaller sample size in this study compared to other studies.

Based on the results of this study, among the studied variables, lack of therapeutic intervention, NVD, male infant, and breastfeeding were associated with increased risk of mother-to-child transmission. Few studies have been conducted on risk factors for vertical transmission of HIV. All of these studies agree on the reducing effect of therapeutic intervention (regardless of type of treatment regimen) on the HIV- infection risk of infants (1, 2, 23, 25). Vertical transmission occurs in three ways: during pregnancy, during labor and delivery, and through breastfeeding (1). As the most common way babies get infected is during the fetus's passage through the birth canal, with cesarean delivery, the possible transmission of this way will be reduced (2). Breastfeeding is also one of the transmission ways of HIV from mother to child, so that the risk of vertical transmission in infants breastfed by infected mothers is 5-20% higher than infants who were not breastfed (1). Accordingly, with cesarean delivery and avoidance of breastfeeding, the mother-to-child transmission of HIV infection can be largely prevented (1-5).

In our study, the male gender was associated with increased risk of infant HIV infection, while in a similar study conducted in France, in contrast to our study, the female gender had a higher risk for HIV infection (25). Also, in studies done in Ethiopia (21) and Brazil (23), living in the village was reported as a risk factor of vertical transmission. However, in our study, no significant differences were observed in the infection rate among infants living in urban and rural areas.

In our study, infants born to mothers in both groups showed no significant differences in prematurity or low birth weight. Some studies have mentioned that using therapeutic intervention in women with HIV can cause

prematurity and low birth weight in their newborns (1, 2). However, as noted, such a relationship was not confirmed in the present study.

It should be mentioned that maternal hypertension and low hemoglobin level, which are known factors influencing the incidence of fetal prematurity and low birth weight (12), had no significant differences in the two groups, and therefore, were not considered as confounding variables.

Conclusion

According to the results of this study, treatment intervention significantly reduces the mother-to-child transmission of HIV. Yet, the rate of vertical transmission in Iran is still higher than the reported rate in developed countries despite treatment interventions. More preventive measures seem to be necessary. Possible reasons for the higher rate of vertical transmission in this study compared to other studies can be the vaginal delivery in mothers with HIV, as well as breastfeeding which were higher in our study population than in other studies.

Financial Support

The present work was supported by a grant from Shiraz University of Medical Sciences. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Conflict of Interest

The authors have no potential conflicts of interest to declare.

Authors' Contributions

All authors have made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data, and have been involved in drafting the manuscript or revising it critically for important intellectual content. All authors read and approved the final manuscript.

Informed consent

Written informed consent was obtained from all participants in the study, and written consent to participate was obtained from the parents of the patient (younger than 16 years of age). Written informed consent for publication of clinical details and images was also obtained from all participants and from the parents of participants under the age of 18.

References

1. World Health Organization. Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants: towards universal access. 2010 [cited 30 January 2012]. http://whqlibdoc.who.int/publications/2010/9789241599818_eng.pdf].
2. World Health Organization. Programmatic update: use of antiretroviral drugs for treating pregnant women and preventing HIV infection in infants. 2012 [cited 6 June 2012]. http://www.who.int/hiv/PMTCT_update.pdf].
3. Ciaranello AL, Perez F, Engelsmann B, Walensky RP, Mushavi A, Rusibamayila A, et al. Cost-effectiveness of World Health Organization 2010 guidelines for prevention of mother-to-child HIV transmission in Zimbabwe. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2013;56(3):430-46.
4. Cotton MF, Marais BJ, Andersson MI, Eley B, Rabie H, Slogrove AL, et al. Minimizing the risk of non-vertical, non-sexual HIV infection in children--beyond mother to child transmission. Journal of the International AIDS Society. 2012;15(2):17377.
5. Ciaranello AL, Perez F, Maruva M, Chu J, Engelsmann B, Keatinge J, et al. WHO 2010 guidelines for prevention of mother-to-child HIV transmission in Zimbabwe: modeling clinical outcomes in infants and mothers. PloS one. 2011;6(6):e20224.
6. Ladner J, Besson MH, Rodrigues M, Sams K, Audureau E, Saba J. Prevention of mother-to-child HIV transmission in resource-limited settings: assessment of 99 Viramune Donation Programmes in 34 countries, 2000-2011. BMC public health. 2013;13:470.
7. Kaida A, Matthews LT, Kanters S, Kabakyenga J, Muzoora C, Mocello AR, et al. Incidence and predictors of pregnancy among a cohort of HIV-positive women initiating antiretroviral therapy in Mbarara, Uganda. PloS one. 2013;8(5):e63411.
8. Gamell A, Letang E, Jullu B, Mwaigomole G, Nyamtema A, Hatz C, et al. Uptake of guidelines on prevention of mother-to-child transmission of HIV in rural Tanzania: time for change. Swiss medical weekly. 2013;143:w13775.
9. Yu L, Li WY, Chen RY, Tang ZR, Pang J, Gui XZ, et al. Pregnancy outcomes and risk factors for low birth weight and preterm delivery among HIV-infected pregnant women in Guangxi, China. Chinese medical journal. 2012;125(3):403-9.
10. Anoje C, Aiyenigba B, Suzuki C, Badru T, Akpoigbe K, Odo M, et al. Reducing mother-to-child transmission of HIV: findings from an early infant diagnosis program in south-south region of Nigeria. BMC public health. 2012;12:184.
11. Johnson LF, Stinson K, Newell ML, Bland RM, Moultrie H, Davies MA, et al. The contribution of maternal HIV seroconversion during late pregnancy and breastfeeding to mother-to-child transmission of HIV. Journal of acquired immune deficiency syndromes (1999). 2012;59(4):417-25.
12. Whitmore SK, Taylor AW, Espinoza L, Shouse RL, Lampe MA, Nesheim S. Correlates of mother-to-child transmission of HIV in the United States and Puerto Rico. Pediatrics. 2012;129(1):e74-81.
13. Merdekios B, Adedimeji AA. Effectiveness of

interventions to prevent mother-to-child transmission of HIV in Southern Ethiopia. International journal of women's health. 2011;3:359-66.
14. Cooper ER, Charurat M, Mofenson L, Hanson IC, Pitt J, Diaz C, et al. Combination antiretroviral strategies for the treatment of pregnant HIV-1-infected women and prevention of perinatal HIV-1 transmission. Journal of acquired immune deficiency syndromes (1999). 2002;29(5):484-94.
15. Lim Y, Kim JY, Rich M, Stulac S, Niyonzima JB, Smith Fawzi MC, et al. Improving prevention of mother-to-child transmission of HIV care and related services in eastern Rwanda. PLoS medicine. 2010;7(7):e1000302.
16. Sedaghat A, Gouya MM, Kamali K, Fahimfar N, Namdari Tabar H, Feiz zadeh A. Estimates the prevalence of HIV using model based design, 2010-2014. Science and Health journal. 2011;5(2):40.
17. WEBDA news agency. Last reported AIDS cases registered in Iran until March of 2013. [cited 28 April 2013]. http://webda.behdasht.gov.ir/index.aspx?siteid=326&pageid=32628&newsview=86794].
18. Lussiana C, Clemente SV, Ghelardi A, Lonardi M, Pulido Tarquino IA, Floridia M. Effectiveness of a prevention of mother-to-child HIV transmission programme in an urban hospital in Angola. PloS one. 2012;7(4):e36381.
19. Chetty T, Knight S, Giddy J, Crankshaw TL, Butler LM, Newell ML. A retrospective study of Human Immunodeficiency Virus transmission, mortality and loss to follow-up among infants in the first 18 months of life in a prevention of mother-to-child transmission programme in an urban hospital in KwaZulu-Natal, South Africa. BMC pediatrics. 2012;12:146.
20. Goswami S, Chakravorty PS. Prevention of Parent to Child Transmission of HIV (PPTCT): An Effort of 4 Years in a Tertiary Centre. Journal of obstetrics and gynaecology of India. 2011;61(4):394-8.
21. Koye DN, Zeleke BM. Mother-to-child transmission of HIV and its predictors among HIV-exposed infants at a PMTCT clinic in northwest Ethiopia. BMC public health. 2013;13(1):398.
22. Neubert J, Pfeffer M, Borkhardt A, Niehues T, Adams O, Bolten M, et al. Risk adapted transmission prophylaxis to prevent vertical HIV-1 transmission: effectiveness and safety of an abbreviated regimen of postnatal oral zidovudine. BMC pregnancy and childbirth. 2013;13:22.
23. Soeiro CM, Miranda AE, Saraceni V, Lucena NO, Talhari S, Ferreira LC. Mother-to-child transmission of HIV infection in Manaus, State of Amazonas, Brazil. Revista da Sociedade Brasileira de Medicina Tropical. 2011;44(5):537-41.
24. Vieira AC, Miranda AE, Vargas PR, Maciel EL. HIV prevalence in pregnant women and vertical transmission in according to socioeconomic status, Southeastern Brazil. Revista de saude publica. 2011;45(4):644-51.
25. Warszawski J, Tubiana R, Le Chenadec J, Blanche S, Teglas JP, Dollfus C, et al. Mother-to-child HIV transmission despite antiretroviral therapy in the ANRS French Perinatal Cohort. AIDS (London, England). 2008;22(2):289-99.
26. Townsend CL, Cortina-Borja M, Peckham CS, de Ruiter A, Lyall H, Tookey PA. Low rates of mother-to-child transmission of HIV following effective pregnancy interventions in the United Kingdom and Ireland, 2000-2006. AIDS (London, England). 2008;22(8):973-81.

Journal of Sciences, Islamic Republic of Iran 36(1): 29 - 38 (2025) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

A New Approach for Normalizing Continuous Data, Applicable in Parametric and Nonparametric Continuous Studies

M.M Saber[1], M. Taghipour[2*], M. Salehi[2], H.M Yousof[3]

[1] *Department of Statistics, Higher Education Center of Eghlid, Eghlid, Islamic Republic of Iran.*
[2] *Department of Statistics, Faculty of Sciences, University of Qom, Qom, Islamic Republic of Iran.*
[3] *Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Benha University, Egypt*

Received: 5 November 2024 / Revised: 31 May 2025 / Accepted: 13 July 2025

Abstract

Satisfying the normality assumption is fundamental to many statistical inferences, as its violation can significantly affect the validity and reliability of conclusions drawn from the data. In this paper, we introduce a novel method for normalizing data that applies to both parametric and non-parametric cases. This method is grounded in a refined version of the empirical distribution function (EDF), which enhances its flexibility and accuracy compared to traditional normalization techniques. By leveraging this new EDF formulation, our approach effectively addresses common issues associated with existing methods, such as sensitivity to outliers and the inability to handle skewed distributions efficiently. A key advantage of our technique is its reversibility, which enables normalized data to be effortlessly transformed back into their original form, thereby preserving the integrity of the raw data for further analysis or interpretation. To demonstrate the efficacy of our method, we evaluate its performance using multiple real-world examples, including datasets related to the COVID-19 pandemic. These datasets, characterized by their complexity and variability, provide a rigorous test of the proposed normalization approach. The results confirm that our method successfully normalizes the data while maintaining their underlying structure and relationships, thus improving the robustness of subsequent statistical analyses. This innovation not only expands the toolkit available for data preprocessing but also enhances the applicability of standard statistical techniques to a broader range of real-life datasets.

Keywords: Normality assumption; Box-Cox transformation; Yeo-Johnson Transformation; Empirical distribution function.

Introduction

Statistical distributions are important in practical fields such as reliability engineering, medicine, and finance. The normal distribution is particularly important among these distributions, especially for classical statistical analysis, including confidence intervals, hypothesis testing, and regression analysis. These

* Corresponding Author: Tel: 09171177616; Email m.taghipour@qom.ac.ir

parametric methods rely heavily on the assumption of normality (1). However, it is important to acknowledge that in certain cases, the central limit theorem cannot be used to assume the normality of the data.

Many researchers often aim to develop new models that retain key features of the original model while also adhering to all necessary assumptions. This can involve techniques such as applying appropriate transformations to the data or filtering out questionable data points that may be regarded as outliers. This approach has been discussed in the works of Thöni (2) and Hoyle (3), among others. Nevertheless, there are techniques available to transform data and align it with the normal distribution. By utilizing such transformations, statistical analysis can be tailored to better suit the specific data and take advantage of the benefits associated with the normal distribution.

Two popular methods for data normalization, the Box-Cox (4, 5) and Yeo-Johnson transformations (6) exist. However, these methods have limitations that may restrict their applicability. For example, the Box-Cox transformation assumes positive and continuous data, with the existence of variance being a crucial condition. Applying this transformation to heavy-tailed data can be problematic and render the results invalid due to the absence of variance in such distributions. Additionally, both Box-Cox and Yeo-Johnson transformations can be substantially influenced by extreme outliers in the dataset, potentially distorting the shape and resulting in distorted normalized data.

Furthermore, while the goal of these transformations is data normalization, the transformed values obtained are not always easy to interpret. Consequently, their use may require additional explanation or conversion back to the original scale for meaningful inference. To overcome these challenges, continued exploration and refinement of normalization techniques are necessary. Identifying and developing new methods that mitigate the limitations of the Box-Cox and Yeo-Johnson transformations would contribute to advances in data normalization and enhance their overall applicability across diverse datasets.

This topic has garnered significant attention from researchers in recent years. For instance, in the paper (8), the Box–Cox transformation a fundamental tool in statistical modeling is comprehensively reviewed and further developed. The authors provide an in-depth exploration of its historical evolution and diverse applications. Additionally, they introduce an extended Yeo-Johnson transformation, which allows for separate power transformations for positive and negative response values. The necessity of this extension is demonstrated through robust data analysis, highlighting its utility in addressing asymmetry and heteroscedasticity in datasets.

Furthermore, Riani et al. (9) propose an automated approach for applying robust versions of the Box–Cox and extended Yeo-Johnson transformations to regression models. This method ensures that the response variable achieves approximate normality, even when it contains both positive and negative values. By incorporating robust statistical techniques, their approach mitigates the influence of outliers and enhances the reliability of model assumptions, thereby improving the overall validity of regression analyses.

In this study, we propose a new transformation for normalizing data that works accurately for all data (positive and negative) with a parametric continuous distribution. This transformation is based on a continuous version of the empirical distribution function (EDF) and can be applied to nonparametric data with the same efficiency. It is also applicable for distributions that do not have a cumulative distribution of closed form. The details of the proposed method, in two cases parametric and nonparametric, are presented in Sections 3 and 4 respectively.

Materials and Methods

In this section, we will examine some techniques for normalizing data, after which we will introduce our proposed method.

1. The Current methods

1.1 The Box-Cox transformation

Tukey (7) introduced a family of power transformations designed to ensure that the transformed values represent a monotonic function of the observations within a suitable range, typically indexed by

$$x_t^{(\lambda)} = \begin{cases} \log(x_t), & \lambda = 0 \\ x_t^{\lambda}, & \lambda \neq 0 \end{cases}$$

for $x_t > 0$.

However, this family was subsequently refined by Box & Cox (1) to address the discontinuity that arises at $\lambda = 0$.

The Box-Cox transformation is a widely used statistical technique for normalizing data. It involves transforming a target variable into a normalized variable using a power transformation. This transformation is controlled by a parameter, the lambda, which is chosen such that it achieves the best approximation to a normal distribution. The Box-Cox method can be applied to various types of data, except negative data, and is commonly used in fields such as finance, economics, and engineering.

To transform a target variable x into a normalized variable w, we use Equation (1), where t represents the period, and λ is a parameter that ranges from -5 to 5.

$$w_t = \begin{cases} \log(x_t), & \lambda = 0 \\ \frac{(x_t^{\lambda}-1)}{\lambda}, & \lambda \neq 0 \end{cases} \quad (1)$$

It is mentioned that the current transformation can be performed on non-time series data as well.

1.2 Yeo-Johnson transformation

Yeo-Johnson transformation is also a statistical technique that normalizes data similar to the Box-Cox transformation. It is an extension of the Box-Cox transformation and can be applied to both positive and negative values of the target variable. The transformation is controlled by a parameter estimated using maximum-likelihood methods to achieve the best approximation of a normal distribution. Yeo-Johnson transformation is commonly used in various fields, including finance, engineering, and social sciences.

The Yeo-Johnson transformation is defined by Equation (2).

$$\psi(\lambda, y) = \begin{cases} \frac{(y+1)^{\lambda}-1}{\lambda}, & \lambda \neq 0, y \geq 0 \\ \log(y+1), & \lambda = 0, y \geq 0 \\ -\frac{(-y+1)^{2-\lambda}-1}{2-\lambda}, & \lambda \neq 2, y < 0 \\ -\log(-y+1), & \lambda = 2, y < 0 \end{cases} \quad (2)$$

If y is strictly positive, then the Yeo-Johnson transformation is the same as the Box-Cox transformation of $(y+1)$.

If y is strictly negative, then the Yeo-Johnson transformation is the Box-Cox transformation of $(-y+1)$ with power $2-\lambda$. For both negative and positive values, the transformation is a mixture of them, using different powers for each.

2. The proposed method for the parametric case

In this section, a link between two continuous random variables through the distribution function method is established. Applying the following theorem, we can derive a transformation for any arbitrary continuous random variable that results in a standard normal distribution.

Theorem 1. Let X and Y be continuous random variables with cumulative distribution functions (CDFs) F_X and F_Y, respectively. If the function h is defined as $h(y) = {F_X}^{-1}(F_Y(y))$; then $h(Y) \stackrel{D}{=} X$.

Proof. The CDF of $h(Y)$ is computed as follows:

$$F_{h(Y)}(x) = P\,(h(Y) \leq x) = P\left({F_X}^{-1}(\mathrm{F_Y(Y)}) \leq x\right) = P\,(F_Y(Y)) \leq F_X(x)),$$

then

$$F_{h(Y)}(x) = P\left(Y \leq {F_Y}^{-1}(F_X(x))\right) = F_Y\left({F_Y}^{-1}(F_X(x))\right) = F_X(x).$$

Since, $F_{h(Y)}(x) = F(x)$ for all x, the proof is complete.

The following corollaries resulted directly from Theorem 1.

Corollary 1. Let X and Y have exponential distribution with parameters λ_1 and λ_2, respectively. Let $F_X$and F_Y be their CDFs. then,

$${F_X}^{-1}(x) = -\frac{1}{\lambda_1}\ln(1-x)$$

and $F_Y(y) = 1 - e^{-\lambda_2 y}$. Therefore,

$${F_X}^{-1}(F_Y(y)) = -\frac{1}{\lambda_1} ln(1 - 1 + e^{-\lambda_2 y}) = \frac{\lambda_2}{\lambda_1} y.$$

Thus, $\frac{\lambda_2}{\lambda_1} Y \stackrel{D}{=} X$.

Corollary 2. Let $X \sim weibull(\alpha, \lambda)$ and $Y \sim logistic(0,1)$, with CDFs F_X and F_Y, respectively. Then,

$${F_X}^{-1}(x) = \left[-\frac{1}{\lambda}\ln(1-x)\right]^{\frac{1}{\alpha}},$$

and $F_Y(y) = \tan^{-1}(y)$.
Therefore,

$${F_X}^{-1}(F_Y(Y))\ \left[-\frac{1}{\lambda}\ln(1-\tan^{-1}(Y))\right]^{\frac{1}{\alpha}} \stackrel{D}{=} X.$$

Now, the basic theorem of this section is stated.

Theorem 2. For any continuous random variable Y with CDF F_Y, the transformation

$$h(Y) = \boldsymbol{\Phi}^{-1}(\mathrm{F_Y}(Y)), \quad (3)$$

has a standard Normal distribution. Here, $\boldsymbol{\Phi}$ is the cdf of standard Normal distribution.

Proof: Substituting X as a standard Normal variable in Theorem 1, proves Theorem 2.

The equation (3) converts data with CDF F_Y to

normalized data. For probable application, the inverse of (3) returns normalized data to the original distribution. In other words,

$$h^{-1}(Z) = {F_Y}^{-1}(\boldsymbol{\Phi}(Z)). \quad (4)$$

This converts the normal random variable Z to a random variable Y with CDF F_Y.

For instance, if $Y \sim \text{Exp}(\lambda)$ then

$$\boldsymbol{\Phi}^{-1}(1 - e^{-\lambda Y}) \sim N(0,1).$$

In practice, we can utilize the "$qnorm$" function directly from the R software as "$\boldsymbol{\Phi}^{-1}$". This means that any dataset $\boldsymbol{y}$ with a cumulative distribution function of F_Y can be transformed into a normally distributed dataset by using $qnorm(F_Y(y))$".

To evaluate the effectiveness of the proposed transformation method, a simulation study is conducted. Suppose that $Y \sim E(\lambda)$ then $\Phi^{-1}(1 - \mathrm{e}^{-\lambda \mathrm{Y}}) \sim \mathrm{N}(0,1)$. Therefore, to transform a sample y drawn from an exponential distribution, we apply the transformation:

" $qnorm(1 - exp\,(-\lambda \boldsymbol{y}))$"

This transformation maps the exponential data into standard normal scores. To assess the performance of this method across different scenarios, various combinations of sample sizes and values of λ are considered. The resulting transformed datasets are then tested for normality using the Kolmogorov–Smirnov and Shapiro–Wilk tests.

The corresponding p-values obtained from these tests are summarized in Table 1. As shown in the table, at the significance level of α=0.05, the original data do not follow a normal distribution, whereas the transformed data satisfy the normality assumption based on both the K-S and S-W tests.

However, after applying a normalizing transformation, normality is satisfied according to both K-S and S-W tests under the level of significance $\alpha = 0.05$. Interestingly, for sample sizes of $n = 5$ and $n = 10$, the K-S test suggests that the original data also follows a normal distribution. However, these results are attributed to the fact that the K-S test is asymptotic and may not perform accurately enough for small sample sizes. Hence, for the two cases, the normality of the original data should only be assessed using the S-W test. After applying the transformation method, the data demonstrates substantial improvements in normality, with K-S values increasing significantly (e.g., from 0.027 to 0.998 for (40,1.5) and from 0.034 to 0.788 for (20,1.5)). Likewise, the S-W test values increase noticeably, with some cases (e.g., (40,1.5) and (5,0.1)) achieving p-values above 0.5, indicating much better adherence to normality. However, for cases with lower λ values and smaller sample sizes (e.g., (10,1.5) and (30,0.4)), normality is not fully achieved, suggesting that the transformation method is more effective for larger datasets and higher rate parameters.

2.1. Application to Skewed data

There are instances where the CDF of the original distribution lacks a closed form. In such cases, we can adapt the equation $F_Y(y) = \boldsymbol{\Phi}(\mathrm{y})$ to normalize the data. It is worth noting that in this context, F_Y (y) represents the CDF of the original distribution. Suppose that Y has a skew-normal distribution denoted by $Y \sim \mathrm{SN}(\alpha)$ for $\alpha \neq 0$. Then,

$$f_Y(y) = 2\phi(y)\boldsymbol{\Phi}(\alpha y),$$

$$F_Y(y) = \int_{-\infty}^{y} 2\phi(w)\boldsymbol{\Phi}(\alpha w)\, dw,$$

where ϕ and $\boldsymbol{\Phi}$ are the standard normal density and distribution function, respectively.

Let $\alpha = 1$, $F_Y(y) = \boldsymbol{\Phi}^2(\mathrm{y})$. Then $\boldsymbol{\Phi}^{-1}(\boldsymbol{\Phi}^2(\mathrm{Y})) \sim N(0,1)$. When $\alpha \neq 1$, $F_Y(y)$ does not a close form. In this case, the transformation is carried out through (4). Therefore,

$$\boldsymbol{\Phi}^{-1}\left(\int_{-\infty}^{y} 2\phi(w)\boldsymbol{\Phi}(\alpha w)\, dw\right) \sim N(0,1).$$

3. The proposed method for nonparametric case

Theorem 2 is a useful tool for normalizing data in various applications. If the distribution of the data is known, this corollary can be implemented directly. However, in cases where the data lacks a parametric

Table 1. Results of Normality Tests for Original and Transformed Exponential Data at α=0.05 Significance Level.

	$(\boldsymbol{n}, \boldsymbol{\lambda})$	$(\mathbf{40}, \mathbf{1.5})$	$(\mathbf{20}, \mathbf{1.5})$	$(\mathbf{10}, \mathbf{1.5})$	$(\mathbf{30}, \mathbf{0.4})$	$(\mathbf{15}, \mathbf{0.4})$	$(\mathbf{5}, \mathbf{0.1})$
Original data	K-S for E(λ)	0.747	0.121	0.719	0.390	0.409	0.479
	K-S	0.027	0.034	0.130	0.031	0.070	0.175
	S-W	0.000	0.000	0.001	0.000	0.000	0.011
Transformed data	K-S	0.998	0.788	0.694	0.540	0.606	0.773
	S-W	0.588	0.578	0.444	0.184	0.200	0.537

distribution, one can utilize the empirical distribution function (EDF) denoted by F_Y^* instead of F_Y in Theorem 2.

Let $Y_{(1)} \leq, Y_{(2)} \leq \cdots \leq Y_{(\mathrm{n})}$ denote the order statistics of the sample. Then,

$$F_Y^*(y) = \begin{cases} 0, & y < Y_{(1)} \\ \frac{i}{n}, & Y_{(i)} \leq y < Y_{(i+1)}; i = 1, \dots, n-1 \\ 1, & y \geq Y_{(n)} \end{cases} \quad (5)$$

The EDF presented in (5) is not continuous and one-to-one, two conditions that are necessary for Theorem 2. To address this issue, we propose a new version of the empirical distribution function called the Generalized Empirical Distribution Function (GEDF), denoted by $F_{new}^*(y)$. The GEDF is defined in the following form and is recommended for use instead.

$$F_{new}^*(y) = \begin{cases} 0, & y < Y_{(1)} - \delta \\ \frac{y - Y_{(1)} + \delta}{n\delta}, & Y_{(1)} - \delta \leq y < Y_{(1)} \\ \frac{1}{n}\left(i + \frac{y - Y_{(i)}}{Y_{(i+1)} - Y_{(i)}}\right), & Y_{(i)} \leq y < Y_{(i+1)}; i = 1, \dots, n-1 \\ 1, & y \geq Y_{(n)} \end{cases} \quad (6)$$

The value of δ can be determined based on the range of data and is typically a small value. Specifically, we use the formula

$$\delta = \min_{i = 1, \dots, n-1} \left[\frac{1}{n^2}\left(Y_{(i+1)} - Y_{(i)}\right)\right]$$

to calculate its value.

The GEDF in (6) is continuous everywhere and is one-to-one for all $y \leq Y_{(n)}$ as well. Its inverse function is given by Equation (7):

$$F_{new}^{*-1}(p) = \begin{cases} Y_{(1)} - \delta(1 - np) & p < \frac{1}{n} \\ Y_{(i)} + (Y_{(i+1)} - Y_{(i)})(np - i) & \frac{i}{n} \leq p \leq \frac{i+1}{n}; i = 1, \dots, n-1 \end{cases} \quad (7)$$

Equations (5) and (6) show that $F_Y{}^*\left(Y_{(i)}\right) = F_{new}^*\left(Y_{(i)}\right) = \frac{i}{n}$. However, in practical statistical applications, we deal with data values $Y_{(i)}$ for $i = 1, \dots, n-1$, and there is no data available in the intervals $(Y_{(i)}, Y_{(i+1)})$ for $i = 1, \dots, n-1$, or below $Y_{(1)}$. Therefore, there is no difference between using (5) or (6) in such cases. However, the key advantage of GEDF is its continuity property, which distinguishes it from EDF. Moreover, as mentioned earlier, GEDF can be used as per Theorem 2 to obtain the desired conversions. To address this issue, we recommend using half ($[\frac{n}{2}]$) for constructing random data using (6) and then normalizing the remaining data with this function. Subsequently, we can swap these two datasets in the next step.

Suppose $y_1, y_2, \dots, y_{\mathrm{n}}$ is a nonparametric data set and let $m_1 = [\frac{n}{2}]$ and $m_2 = n - m_1$. Choose $Y_1^{(1)}, Y_2^{(1)}, \dots, Y_{m_1}^{(1)}$ randomly from $y_1, y_2, \dots, y_{\mathrm{n}}$ and the name remained data as $Y_1^{(2)}, Y_2^{(2)}, \dots, Y_{m_2}^{(2)}$. The two following transformations convert $y_1, y_2, \dots, y_{\mathrm{n}}$ to normalized data.

$$\delta^{(1)} = \min_{i = 1, \dots, m_1 - 1} \left(\frac{Y_{(i+1)}^{(1)} - Y_{(i)}^{(1)}}{m_1^2}\right)$$

$$\delta^{(2)} = \min_{i = 1, \dots, m_2 - 1} \left(\frac{Y_{(i+1)}^{(2)} - Y_{(i)}^{(2)}}{m_2^2}\right)$$

For $Y_1^{(1)}, \dots, Y_{m_1}^{(1)}$, $h(y)$ is used as follows

$$h(y) = \begin{cases} -3.5, & y < Y_{(1)}^{(2)} - \delta^{(2)} \\ \boldsymbol{\Phi}^{-1}\left(\frac{y - Y_{(1)}^{(2)} + \delta^{(2)}}{m_2\,\delta^{(2)}}\right), & Y_{(1)}^{(2)} - \delta^{(2)} \leq y \leq Y_{(1)}^{(2)} \\ \boldsymbol{\Phi}^{-1}\left(\frac{i}{m_2} + \frac{y - Y_{(i)}^{(2)}}{m_2\left[Y_{(i+1)}^{(2)} - Y_{(i)}^{(2)}\right]}\right), & Y_{(i)}^{(2)} \leq y \leq Y_{(i+1)}^{(2)}; i = 1, \dots, m_2 - 1 \\ 3.5, & Y_{(m_2)}^{(2)} < y \end{cases} \quad (8)$$

and for $Y_1^{(2)}, \dots, Y_{m_2}^{(2)}$, $h(y)$ is applied in the following

$$h(y) = \begin{cases} -3.5, & y < Y_{(1)}^{(1)} - \delta^{(1)} \\ \boldsymbol{\Phi}^{-1}\left(\frac{y - Y_{(1)}^{(1)} + \delta^{(1)}}{m_1\,\delta^{(1)}}\right), & Y_{(1)}^{(1)} - \delta^{(1)} \leq y \leq Y_{(1)}^{(1)} \\ \boldsymbol{\Phi}^{-1}\left(\frac{i}{m_1} + \frac{y - Y_{(i)}^{(1)}}{m_1\left[Y_{(i+1)}^{(1)} - Y_{(i)}^{(1)}\right]}\right), & Y_{(i)}^{(1)} \leq y \leq Y_{(i+1)}^{(1)}; i = 1, \dots, m_1 - 1 \\ 3.5, & Y_{(m_1)}^{(1)} < y \end{cases} \quad (9)$$

The inverse conversion for returning normalized data

to the original is

$$h^{-1}(z) = \begin{cases} Y_{(1)} - \delta\big(1 - n\,\Phi(z)\big), & \Phi(z) < \frac{1}{n} \\ Y_{(i)} + \big(Y_{(i+1)} - Y_{(i)}\big)(n\Phi(z) - i), & \frac{i}{n} \le \Phi(z) \le \frac{i+1}{n}; i = 1, \dots, n-1 \end{cases} \quad (10)$$

where δ is defined as

$$\delta = \min_{i = 1, \dots, n-1} \left(\frac{Y_{(i+1)} - Y_{(i)}}{n^2}\right).$$

4. Application to read data sets

In this section, we demonstrate proficiency of the proposed method in two real data sets.

4.1. Confirmed case rate of the COVID-19 data

This data demonstrates the effectiveness of normalizing transformations for nonparametric data using several real-world datasets. Specifically, we analyze the confirmed case rate (CCR) of the COVID-19 virus across ten different countries: Canada, France, Germany, Iran, Iraq, Italy, Mexico, Netherlands, Turkey, and the UK. These datasets were sourced from publicly available data provided by the World Health Organization (WHO) at https://covid19.who.int/

The CCR data exhibit significant variability and skewness, making them ideal candidates for normalization techniques. To address this, we applied Equations (9) and (10), which represent mathematical transformations designed to reduce skewness and approximate a normal distribution. After applying these transformations, we evaluated the normalized data using two widely used statistical tests: The Kolmogorov-Smirnov (K-S) test and the Shapiro-Wilk (S-W) test. The K-S test assesses the goodness-of-fit between the empirical distribution of the data and a theoretical normal distribution, while the S-W test is particularly sensitive to deviations from normality in smaller sample sizes. The results of both tests are reported in Table 2, where it is evident that the transformations significantly improve the normality of the datasets. This improvement is reflected in the p-values of the tests, which indicate a closer adherence.

The table clearly shows that none of the original datasets are normally distributed, while all datasets normalized using nonparametric transformations have a normal distribution.

To have a comparison with previous methods of normalization, data also were transformed by Box-Cox transformation with the best value of λ. This work is done by function "boxcox" in package "MASS". The results demonstrate that none of these transformed data have a normal distribution regarding the K-S test. However, regarding the S-W test, the transformed data of 4 countries have a normal distribution. This certifies the better performance of our method concerning, for, to Box-Cox transformation.

The results in Table 2 indicate that the original COVID-19 CCR data strongly deviates from normality, as evidenced by extremely small p-values from both the K-S and S-W tests. This suggests that the data distribution is significantly non-normal across all countries. Upon applying the Box-Cox transformation, slight improvements are observed in normality, particularly in the S-W test for some countries (e.g., Canada: 0.2695, Mexico: 0.7005, and the UK: 0.833). However, many values remain well below conventional thresholds for normality ($p > 0.05$), indicating that the

Table 2. Normality tests for original and transformed CCR of COVID-19 data.

	Country	**Canada**	**France**	**Germany**	**Iran**	**Iraq**
Original data	K-S	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$
	S-W	0.0073	$< 2\times 10^{-16}$	2×10^{-15}	4×10^{-16}	3.9×10^{-7}
Transformed data by Box-Cox	K-S	$< 2.2\times 10^{-16}$	$< 2.2\times 10^{-16}$	$< 2.2\times 10^{-16}$	$< 2.2\times 10^{-16}$	$< 2.2\times 10^{-16}$
	S-W	0.2695	1.49×10^{-5}	8.20×10^{-11}	1.89×10^{-5}	8.59×10^{-7}
Transformed data by method of paper	K-S	0.4992	0.9991	0.3439	0.9642	0.786
	S-W	0.1385	0.8113	0.8621	0.2411	0.4
	Country	**Italy**	**Mexico**	**Netherlands**	**Turkey**	**UK**
Original data	K-S	$<7\times 10^{-16}$	$< 2\times 10^{-16}$	1.2×10^{-15}	$< 2\times 10^{-16}$	4.6×10^{-7}
	S-W	0.0047	1.6×10^{-6}	0.0433	$< 2\times 10^{-16}$	0.8188
Transformed data by Box-Cox	K-S	$<7\times 10^{-15}$	$< 2.2\times 10^{-16}$	1.221×10^{-15}	$< 2.2\times 10^{-16}$	2.49×10^{-13}
	S-W	0.0422	0.7005	0.8813	1.30×10^{-9}	0.833
Transformed data by method of paper	K-S	0.9635	0.8221	0.9409	0.6051	0.7644
	S-W	0.2861	0.5015	0.2793	0.2368	0.2571

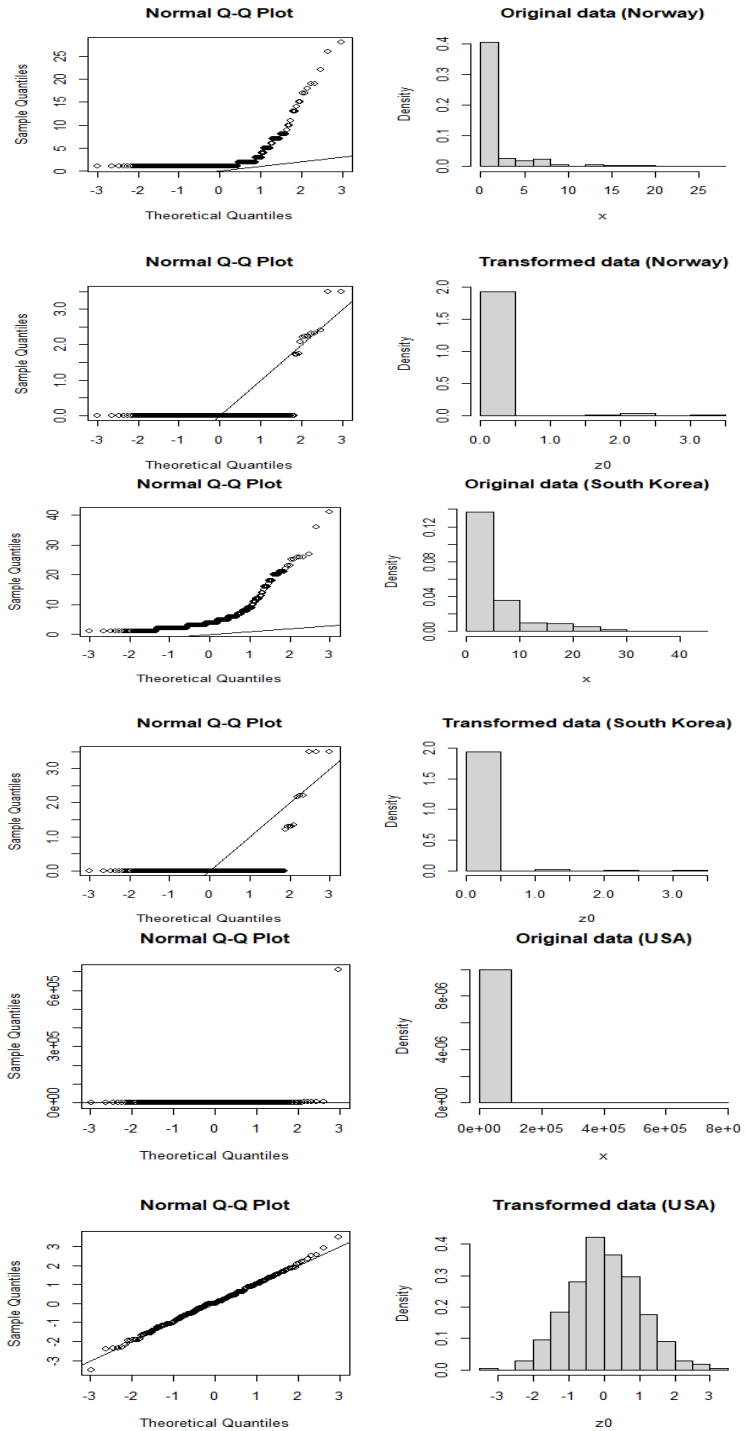

Figure 1. "Effect of Proposed Transformations on Normality: Q-Q Plots and Density Curves for Norway, South Korea, and the USA"

Box-Cox transformation alone is insufficient for achieving a normal distribution in many cases.

Conversely, the transformation method proposed in the paper demonstrates a substantial improvement in normality. K-S test results show much higher p-values (closer to 1), indicating that the transformed distributions better fit a normal distribution. Likewise, the S-W test results show a significant shift towards normality, with values such as 0.8113 (France), 0.8621 (Germany), and 0.5015 (Mexico) reflecting stronger normality characteristics than the Box-Cox transformation. This suggests that the proposed transformation method is more effective at normalizing the data, making it a preferable approach for statistical modeling.

Figure 1 illustrates the effects of the proposed transformation on datasets from Norway, South Korea, and the United States. The original data demonstrated significant non-normality, primarily characterized by right skewness. Following the transformation, the data aligned much more closely with a normal distribution, as evidenced by decreased skewness and the emergence of more symmetrical, bell-shaped density curves. Although some minor deviations from normality persist in certain cases, the transformation markedly enhances the data's suitability for parametric analyses. Each subplot includes a Normal Q-Q plot and a density curve, providing clear visual confirmation of the transformation's success in improving distributional symmetry and overall normality.

Figure 2 displays a grayscale heatmap of log10-transformed p-values from normality tests, illustrating how well COVID-19 CCR data conform to a normal distribution across different countries and transformation techniques. In this visualization, darker shades represent stronger evidence of deviation from normality (lower p-values), while lighter shades indicate data closer to a normal distribution (higher p-values). For example, the original data for Canada are depicted in very dark hues, reflecting extremely low p-values (<1e-16) and a significant departure from normality. In contrast, after applying the proposed transformation, the color intensity becomes noticeably lighter, with p-values exceeding 0.3, demonstrating a marked improvement in normality fit. This comparison underscores the effectiveness of the proposed method, which outperforms both the untransformed data and the Box-Cox transformation in achieving normality.

4.2. The mortality rate and confirmed case rate of COVID-19 data

It is essential to clarify that this article focuses exclusively on the normalization of continuous or non-discrete data. As such, the methodology proposed herein may not be suitable for normalizing discrete data, a point we will further explore in this section through a practical situation that highlights how aggregated data can sometimes be treated as discrete.

We examine the mortality rate (MR) and confirmed case rate (CCR) of COVID-19 across three countries: the USA, South Korea, and Norway. The relevant data can be downloaded from https://covid19.who.int/

In this case, the MR and CCR data for South Korea and Norway contain many zeros and are thus approximately discrete, while the same variables for the USA are continuous. As shown in Table 3, both the K-S and S-W tests confirm that none of the six datasets has a normal distribution. Although the transformed datasets for the USA are normally distributed, the data

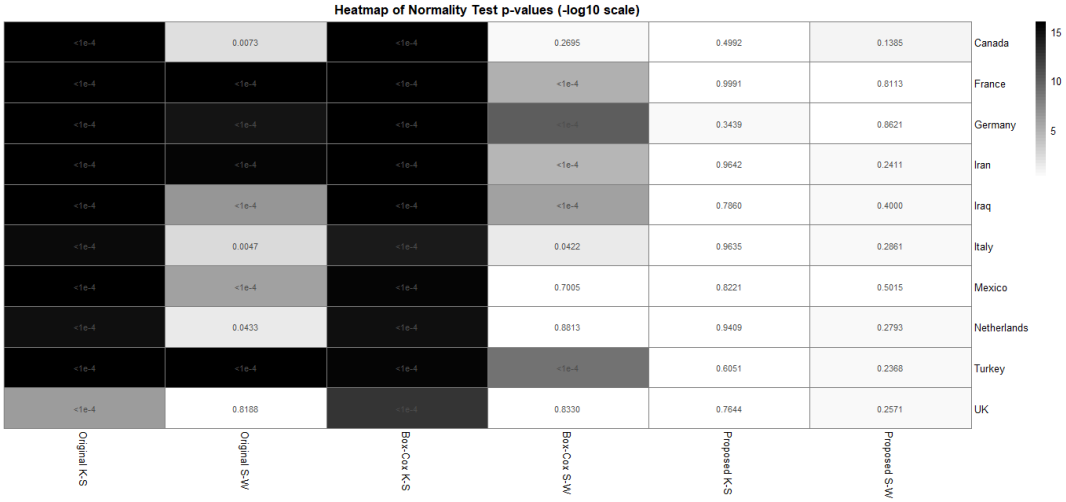

Figure 2 "Comparison of Normality Fit for COVID-19 CCR Data Across Countries: Grayscale Heatmap of Log10-Transformed p-Values"

Table 3. Normality tests for original and transformed MR and CCR of COVID-19 data**.**

	Country	**USA**		**South Korea**		**Norway**	
	variable	**MR**	**CCR**	**MR**	**CCR**	**MR**	**CCR**
Original data	K-S	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$
	S-W	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	8×10^{-14}	$< 2\times 10^{-16}$	8×10^{-14}	$< 2\times 10^{-16}$
Transformed data by Box-Cox	K-S	$< 2.2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$
	S-W	3.23×10^{-6}	$< 2\times 10^{-16}$	4.87×10^{-7}	1.12×10^{-10}	$< 2\times 10^{-16}$	$< 2\times 10^{-16}$
Transformed data by method of paper	K-S	0.9912	0.431	8.69×10^{-8}	$< 2\times 10^{-16}$	3.5×10^{-5}	$< 2\times 10^{-16}$
	S-W	0.7046	0.5023	3.27×10^{-8}	$< 2\times 10^{-16}$	2.9×10^{-6}	$< 2\times 10^{-16}$

for South Korea and Norway do not have the normal distribution yet. The results of the Box-Cox transformation demonstrate that none of these transformed data have a normal distribution regarding the K-S and S-W tests. Therefore, our method has better performance of Box-Cox transformation.

Table 3 presents normality test results for the Mortality Rate (MR) and Case Fatality Rate (CCR) of COVID-19 data across the USA, South Korea, and Norway. The original data exhibits strong deviations from normality, as indicated by extremely small p-values in both the Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests, confirming significant non-normality. Applying the Box-Cox transformation yields minor improvements in a few cases (e.g., the S-W test for MR in the USA: 3.23×10^{-6}), but in general, normality is not achieved, as many values remain exceptionally low. This suggests that the Box-Cox transformation is insufficient for properly normalizing the data. On the other hand, the proposed transformation method in the paper significantly enhances normality for MR in the USA, where K-S and S-W tests return p-values of 0.9912 and 0.7046, respectively. However, for CCR in South Korea and Norway, the transformation does not substantially improve normality, with many p-values remaining near zero. This indicates that while the method is effective in certain cases, its performance varies across datasets and variables, necessitating further investigation.

Results

The normality of data and its distribution is crucial for many statistical analyses, as it simplifies the statistical modeling process. In this article, recognizing the importance of normalization conversions and their application in various scientific fields, we have introduced a new and straightforward normalization conversion method for both parametric and non-parametric continuous data. Our approach outperforms conventional methods such as the Yeo-Johnson transformation and Box-Cox transformation and overcomes their limitations. Additionally, we have introduced a new empirical distribution function that allows the proposed conversion to be used for non-parametric data. Unlike conventional empirical distribution functions, this new definition is continuous, enabling us to normalize non-parametric data.

Conclusion

To further strengthen the study, additional validation should be conducted by comparing the proposed transformation with other techniques such as Yeo-Johnson and log transformations, along with supplementary normality tests like the Anderson-Darling test. Visualizing the distributions through histograms, density plots, and Q-Q plots would provide intuitive confirmation of normality improvements. Moreover, evaluating the impact of normalization on downstream statistical analyses, such as regression modeling and hypothesis testing, would demonstrate its practical benefits. Expanding the application to other epidemiological datasets and periods would help assess its generalizability. Finally, a deeper theoretical discussion on why the proposed transformation outperforms Box-Cox, particularly in terms of skewness and kurtosis reduction, would provide a stronger mathematical justification for its effectiveness.

Moreover, to further validate the findings, the proposed transformation should be compared with alternative techniques such as Yeo-Johnson or log transformations, and additional normality tests like the Anderson-Darling test should be conducted. Visualizing the distributions through histograms, Q-Q plots, and density plots would provide clearer insights into the effectiveness of each transformation. Evaluating the impact of normality improvements on subsequent statistical analyses, such as predictive modeling and hypothesis testing, would help establish its practical benefits. Additionally, applying the method to a broader range of countries and datasets would assess its generalizability. Finally, a theoretical discussion on why

the transformation is more effective for MR than CCR, possibly exploring the influence of skewness and kurtosis, would strengthen the study's conclusions.

References

1. Graybill FA. The Theory and Applications of the Linear Model (London, Duxbury Press). 1976.
2. Thöni H. A table for estimating the mean of a lognormal distribution. Journal of the American Statistical Association. 1969 Jun 1;64(326):632-6
3. Hoyle MH. Transformations: An introduction and a bibliography. The International Statistical Review. 1973; 41(203-223).
4. Box GEP and Cox, DR. An analysis of transformations. Journal of the Royal Statistical Society. 1964 B; 26(211-234).
5. Sakia RM. The Box-Cox transformation technique: A review. The Statistician. 1992; 41(169-178).
6. Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. Biometrika. 2000 Dec 1;87(4):954-9.
7. Tukey JW. The comparative anatomy of transformations. Annals of Mathematical Statistics 1957; 28(602-632).
8. Atkinson AC, Riani M, Corbellini A. The box–cox transformation: Review and extensions. Statistical Science. 2021;36(2):239–55.
9. Riani M, Atkinson AC, Corbellini A. Automatic robust Box–Cox and extended Yeo–Johnson transformations in regression. Statistical Methods & Applications. 2023;32(1):75-102.

Journal of Sciences, Islamic Republic of Iran 36(1): 39 - 49 (2025) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Hybrid Prediction Models for Suicide Mortality Levels in Iranian Provinces: Spatial Econometrics vs. Random Forests

M. A. Farzammehr[1*] and Geoffrey J. McLachlan[2]

[1] *Judiciary Research Institute, Tehran, Islamic Republic of Iran*
[2] *Department of Mathematics, University of Queensland, Brisbane, Australia*

Received: 1 July 2025 / Revised: 25 September 2025 / Accepted: 15 November 2025

Abstract

With the growing utilization of advanced machine-learning techniques, such as random forests, understanding the significance of spatial factors within these models is increasingly imperative. This study proposes a novel approach to develop spatially explicit classification random forest models by integrating spatially lagged variables, mirroring various spatial panel data econometric specifications. We assess the comparative performance of these models against traditional spatial and non-spatial regression methods to predict suicide mortality rates across 31 provinces in Iran, utilizing data from 2011 to 2021. Results reveal that the spatial random forest model, incorporating spatial lag parameters, achieves a remarkable accuracy of 89.19% in predicting suicide mortality levels, surpassing traditional spatial econometric models (46.51%) and non-spatial random forest models (27.03%). While highlighting the effectiveness of spatial random forest models with spatial lag parameters, this study also recognizes the continued relevance of traditional spatial econometric models in predicting suicide mortality rates. These findings offer valuable insights into the interplay between spatial considerations and predictive modeling, providing essential guidance for researchers in selecting appropriate models for spatial data analysis.

Keywords: Spatial Econometrics models; Machine-Learning Techniques; Random Forests; Iran.

Introduction

Suicide mortality is a pressing public health concern, with devastating implications for individuals, families, and communities. In this profound challenge, accurate predictive models are essential to inform timely interventions and targeted preventive measures. The field of data science has responded to this need with the emergence of new machine-learning techniques, exemplified by the versatile random forest algorithm, capable of handling large and complex datasets (1, 2, 3). These techniques offer promising avenues for predictive modeling and have garnered significant attention in various domains.

However, as the scope of data science expands, so does the recognition of the critical role that spatial considerations play in the accuracy and applicability of predictive models. Spatial data, characterized by geographic and interdependencies, adds layer of complexity to the modeling process. Understanding the spatial context is pivotal for unraveling underlying patterns and relationships that contribute to suicide

* Corresponding Author: Tel:+989133051096; Email: m.farzammehr@jri.ac.ir

mortality rates, as well as for identifying spatially varying risk factors and vulnerability (4, 5).

Integrating spatial considerations into machine-learning models is crucial for several reasons (6, 7). First, traditional machine-learning techniques, including random forests, have been predominantly developed for predictive applications, often leaving researchers with limited insights into the underlying explanatory relationships between variables. Consequently, these models are often regarded as "black boxes" that limit both the interpretation and the generalization of results, especially in spatial contexts where understanding spatial interactions is essential.

Second, the spatial dimensions of the data demand specialized attention. Spatial econometrics, a well-established branch of spatial statistics, has long been utilized to analyze spatially explicit data, accounting for spatial autocorrelation, spatial heterogeneity, and spatial dependence. Nevertheless, the recent surge of interest in machine-learning techniques has left open questions about the effectiveness of spatial econometrics compared to machine-learning models such as random forests.

Despite the increasing application of machine-learning techniques and spatial econometrics as separate approaches (8, 9, 10), there is still a lack of studies that systematically integrate the two for the prediction of suicide mortality, especially in the context of Iranian provinces. To address this gap, this paper introduces a hybrid framework that incorporates spatially lagged variables into random forest models, enabling a direct comparison with traditional spatial econometric approaches and advancing a novel methodology for spatially explicit predictive modeling. To the best of our knowledge, this is the first study to systematically integrate spatial econometric lags into random forests for predicting suicide mortality in Iranian provinces.

This paper addresses these crucial issues by presenting a novel approach that bridges the gap between spatial econometrics and random forests. We propose constructing spatially explicit classification Random Forest models by thoughtfully incorporating spatially lagged variables. By emulation, we mean that the inclusion of different types of spatial lags—such as lagged dependent variables and lagged explanatory variables—allows the random forest to approximate the relationships captured by spatial panel data econometric models (e.g., SLM, SDM), without formally distinguishing between them in a strict econometric sense. This design enables us to test the comparative performance of these hybrid models against both traditional spatial econometric techniques and non-spatial regression methods. The focus is on predicting suicide mortality rates across 31 provinces in Iran over eleven years (2011–2021), utilizing a comprehensive spatial panel dataset.

This study addresses fundamental questions at the forefront of predictive modeling research. Our focus centers on comparing two distinct approaches: the predictive machine-learning models, exemplified by the random forest algorithm, and the traditional spatial econometric techniques. Moreover, we endeavor to ascertain whether spatially explicit random forest models possess the potential to outperform standard non-spatial random forest models when applied to the crucial task of predicting suicide mortality rates. By unraveling the answers to these pivotal questions, our research aims to provide valuable insights into the interplay between machine-learning techniques and spatial econometrics, opening new avenues for improved predictive modeling in spatially explicit contexts.

By assessing the predictive accuracy and interpretability of these models, we provide findings that shed light on the relationship between spatial considerations and predictive modeling. We aim to provide essential guidance for researchers in selecting appropriate models for spatial data analysis, particularly in the context of suicide mortality prediction. Ultimately, this study advances the expanding literature on spatially explicit machine learning methods while providing direction for future research on their spatial implications and potential applications.

As the importance of spatially aware data analysis grows, our study aims to inform suicide prevention strategies and underscore the significance of embracing spatial considerations in predictive modeling endeavors.

The remainder of this paper proceeds as follows. Section 2 reviews the relevant literature on spatial econometrics and machine learning applications. Section 3 describes the data collection and preparation process. Section 4 presents the estimation methodology, including integrating spatially lagged variables into random forest models. Section 5 reports the results, including model performance comparisons and interpretation of key findings. Finally, Section 6 concludes the paper, highlighting the main contributions, limitations, and directions for future research.

Literature Review

In machine learning, which finds diverse applications across various domains, including crime analysis, it is of utmost importance to establish foundational definitions and distinctions. Machine-learning methods, a subset of artificial intelligence (AI), are designed to learn from data and effectively predict outcomes for new, previously unseen data (11). These methods fall into two main categories: supervised and unsupervised techniques.

Supervised algorithms operate by learning the relationship between independent and dependent variables, which they then use to predict outcomes for new instances. This type of prediction enables the algorithm to generalize its learning from the training data to unseen data. Within the realm of supervised learning, we encounter a diverse array of techniques, ranging from conventional statistical methods such as linear and logistic regression to more advanced and sophisticated approaches such as support vector machines (SVM), decision trees, random forests, and neural networks (12).

On the other hand, unsupervised algorithms generate predictions, often in the form of classifications, solely based on the underlying relationships among the provided independent variables. Examples of such techniques include clustering methods such as k-means and dimensionality reduction techniques such as principal component analysis (PCA) (12). These approaches are valuable for tasks where labeled data is absent, as the determining factor for their application is the absence of labeled data rather than clear relationships between variables.

This categorization of machine-learning methods lays a solid foundation for understanding their respective applications and potential advantages in crime analysis and other domains. Combining traditional statistical techniques and cutting-edge machine-learning algorithms allows researchers and practitioners to tackle complex problems and make accurate predictions, thereby contributing to advancements in various fields of study.

While these distinctions are well-suited for predictive applications, they can perplex those accustomed to using these statistical techniques for explanatory purposes. Additionally, some of the newer machine-learning algorithms, such as decision trees, random forests, and neural networks, do not explicitly measure the relationship between independent and dependent variables, leading to criticism and being labeled as "black box" methods. Consequently, the more critical distinction lies between primarily predictive machine-learning techniques, such as random forests, which do not produce coefficients or explicit relationship measures between independent and dependent variables, and the more traditional explanatory machine-learning techniques, such as linear regression, that are familiar with generating such measures. However, it is worth noting that even predictive models such as random forests can still provide insights into explanatory features, as demonstrated in the application within this paper.

Moreover, alongside applying existing predictive machine-learning methods to Spatial Criminology inquiries, scholars have recognized the importance of developing spatially explicit predictive models and methods (7, 13, 14). Singleton and Arribas-Bel (2019) succinctly emphasize that one of the most promising areas for Geographic Data Science to enhance core data science techniques is explicitly incorporating spatial information to improve the performance of predictive machine-learning models. Notably, a limited number of existing studies that have created spatially explicit prediction frameworks reveal that these methods tend to outperform non-spatial methods when applied to spatial datasets (13, 15, 16, 17).

To gain deeper insights into the role of space in predictive machine-learning models, this paper systematically examines various spatial lags in a random forest model, compares random forest models to more conventional spatial econometric models, and evaluates the factors that contribute to enhanced predictive performance for either random forests or spatial econometric models. In the context of our research, we draw from these prior investigations and aim to shed light on the interplay between spatial considerations and predictive modeling. By doing so, our study will offer valuable guidance to researchers seeking to harness the potential of machine learning for spatial data analysis, specifically in predicting suicide mortality rates across provinces in Iran. The outcomes of this investigation will have implications for the development of improved public health policies and interventions in the region, thereby contributing to the broader body of knowledge in spatial data analysis and machine learning.

Materials and Methods

Data Collection and Preparation

Suicide mortality data in Iran come from reports published by the Iranian Forensic Medicine Organization (IFMO), which operates under the Iranian Judicial Authority (5). The IFMO keeps a comprehensive suicide registry and conducts autopsies for all recorded suicide cases (4). Suicide rates per 100k individuals for each province and socio-demographic and economic data for all 31 provinces were obtained from the Statistical Center of Iran.

The data encompass 2011 to 2021, representing the most recent available data. The variables collected include unemployment rate (X1), labor force participation rate (X2), natural log of population aged 15 and over (X3), consumer price index or CPI (X4), literacy rate among individuals aged six and older (X5), and natural log of gross domestic product or GDP (X6). These variables will be used to identify factors influencing suicide mortality rates per 100k populations

(y) using an econometric model that considers spatial correlations. The explanatory variables X1 to X6 are available for each of the 31 provinces, allowing for a comprehensive, province-specific analysis and modeling.

The study employs panel data covering 31 provinces over 11 years. This panel data structure enables the analysis of individual provinces over time and cross-sectional variations among provinces. It provides a robust framework for examining spatial and temporal dynamics, enhancing the understanding of the factors influencing suicide mortality rates in Iran.

Table 1 provides a descriptive overview of the independent variables used in the study. The dataset used in our experiment comprises a total of 341 instances. To enable prediction, a new attribute called 'suicide category' was added to the dataset. Converting the target (class) attribute to a nominal type was necessary for prediction purposes. As all the original characteristics in the dataset were numeric, the new nominal attribute 'suicide category' was introduced to facilitate prediction.

The 'suicide category' attribute categorizes cases based on the percentage of suicide mortality. Instances with 'suicide rates per 100k individuals' below 33% fall into the 'Low' category in the 'suicide category' attribute.

Those with 'suicide rates per 100k individuals' equal to or greater than 33% but less than 66% were labeled as 'Medium'. Finally, instances with 'suicide rates per 100k individuals' equal to or greater than 66% were categorized as 'High' in the 'suicide category'. The values for the 'suicide category' attribute were meticulously calculated for all 341 instances and cross-checked multiple times by all authors to ensure accuracy and eliminate any potential errors.

Since the analysis is conducted at the provincial level, spatial proximity may induce correlations among the residuals. The incorporation of spatial considerations is crucial to account for potential spatial dependencies and patterns in the data, allowing for a more accurate and comprehensive analysis of suicide mortality across the Iranian provinces. Spatial autocorrelation analysis is a vital aspect of understanding the patterns and dynamics of suicide mortality rates across geographical regions.

To explore spatial dependencies of the 31 Iranian provinces, we employ the queen-contiguity weight matrix (18). This matrix is specifically tailored to the polygonal nature of our data and provides a comprehensive definition of neighborhood relationships. It captures spatial connectivity based on shared vertices between neighboring spatial units. In our dataset, provinces can be considered as polygonal areas with irregular shapes, and the queen-contiguity matrix efficiently captures spatial relationships between these areas. Specifically, for each province, neighboring provinces that share at least one common vertex are considered connected. This definition ensures that spatial interactions are accounted for not only through shared boundaries but also through shared vertices, providing a more complete representation of spatial relationships.

For transparency and reproducibility, we explicitly document the rationale for our methodological decisions. First, the queen-contiguity specification captures spatial interactions among provinces effectively and follows best practices in regional health research (18, 19), ensuring comparability with prior studies. Second, random forest hyperparameters were systematically optimized through repeated cross-validation—rather than arbitrarily selected—focusing on the number of trees, maximum tree depth, and minimum sample size at terminal nodes (2, 20). This process balanced predictive accuracy with the need to mitigate overfitting. Finally, during data preprocessing, spatially lagged variables derived from econometric specifications were generated and incorporated directly into the random forest training process (6, 21). This integration preserved the predictive flexibility of machine learning while explicitly modeling spatial spillover effects. By documenting these methodological choices and their rationale, we enhance the reproducibility of the study and provide a transparent roadmap for future research using similar spatial panel datasets.

In spatial econometrics, selecting the appropriate model specification is crucial, and an empirical specification test is a valuable tool for this purpose. The specificity-to-generality approach is commonly employed, starting with a basic non-spatial model and

Table 1. Overview of Descriptive Statistics for Dependent and Independent Variables

Variable	Minimum	Mean	Maximum	SD	Skewness	Kurtosis
X1	5.80	11.20	21.70	3.07	0.90	4.16
X2	32.30	41.38	50.20	3.62	-0.28	3.01
X3	12.38	14.13	16.46	0.77	0.12	2.63
X4	39.20	139.40	401.00	97.43	1.18	3.26
X5	70.80	84.49	92.90	4.36	-0.62	3.52
X6	10.30	11.77	14.39	0.85	0.59	3.23

testing for misspecifications arising from omitted autocorrelation. To address this issue, Anselin et al. (1996) introduced the Lagrange multiplier (LM) tests, which offer robustness against alternative sources of spatial dependence.

Spatial dependencies refer to the interdependence and interaction of observations in space, where the values of neighboring observations can influence each other. Neglecting this spatial autocorrelation can lead to biased and inefficient parameter estimates in the model. The LM test focuses on spatial autocorrelation in the residuals of a non-spatial econometric model, such as the Ordinary Least Squares (OLS) model. Its fundamental idea is to examine whether the residuals exhibit a systematic spatial pattern. A significant result in the LM test indicates that the model is mis-specified and fails to account for spatial dependencies in the data.

By conducting the LM test, we can identify the presence and extent of spatial autocorrelation. Significant spatial autocorrelation indicates that a spatial econometric model is more appropriate for the data. This implies that neighboring observations interact, so the model explicitly considers their spatial relationships. Including of geographical attributes in the model is one way to address significant and positive spatial autocorrelation.

Table 2 presents Lagrange multiplier diagnostics for spatial dependence. The Lagrange Multiplier test for error spatial dependence (LMerr) examines spatial autocorrelation in the residuals of a non-spatial econometric model. In contrast the Lagrange Multiplier test for lag spatial dependence (LMlag) focuses on spatial autocorrelation in the dependent variable. The Robust Lagrange Multiplier test for error spatial dependence (RLMerr) and the Robust Lagrange Multiplier test for lag spatial dependence (RLMlag) offer robust versions of these diagnostics, enhancing reliability in detecting spatial autocorrelation. Additionally, the Spatial Autoregressive Moving Average (SARMA) model, as discussed in Haining (2003), incorporates spatial autocorrelation, spatial heterogeneity, and spatial dependence in the modeling process. By applying these Lagrange multiplier tests, we can detect and correct spatial autocorrelation in the data, thereby ensuring the validity of our spatial econometric models (22, 23).

The test results reveal that LMerr, LMlag, and RLMerr tests show highly significant P-values (below the 0.05 significance level), indicating strong evidence of spatial autocorrelation in the residuals and the dependent variable. Additionally, RLMlag exhibits a P-value smaller than 0.05, suggesting the presence of spatial autocorrelation in the dependent variable. The SARMA model further supports the significant spatial autocorrelation in the data.

In conclusion, the LM tests are crucial in identifying spatial autocorrelation and supporting the inclusion of spatial lag terms in the econometric model. By explicitly considering spatial relationships among observations, we can ensure a more accurate and robust analysis, which is vital for understanding and addressing spatial patterns in the data.

Estimation Methodology

This study aims to compare the predictive accuracy of traditional spatial econometric models with random forest models for predicting suicide mortality levels in the provinces of Iran. To achieve this, we employed eight distinct model specifications: spatial lag (autoregressive) (SAR), spatially lagged **X** (SLX), spatial Durbin (SDM), random forests (RF), random forests with the spatial lag of **y** included (RFSAR), random forests with spatial lags of both **X** and **y** included (RFSDM), and random forests with only the spatial lag of **X** included (RFSLX). Each model provides distinctive insight into the relationship between spatial dependencies and suicide mortality. In the subsequent sections, we delve into the comprehensive estimation details of these models to uncover their strengths and limitations in predicting suicide mortality levels.

Spatial Econometric Models: Spatial econometric models capture spatial dependence in the data-generating process, recognizing that objects in proximity exhibit stronger relationships than those farther apart, a concept known as Tobler's First Law of Geography (24). These models extend the Ordinary Least Squares (OLS) approach by incorporating a spatial weights matrix (**W**) that represents the spatial relationships between observations directly into the model estimation process

Table 2. Lagrange Multiplier Diagnostics for Spatial Dependence

Test	Statistic	P-value
LMerr	417.06	< 2.2e-16
LMlag	247.38	< 2.2e-16
RLMerr	176.3	< 2.2e-16
RLMlag	6.6259	0.01005
SARMA	423.68	< 2.2e-16

(19, 22). The placement of the spatial weights matrix in the standard linear regression equation can vary based on theoretical or data-driven considerations, resulting in different model specifications that account for spatial autocorrelation.

The most basic linear regression (OLS) specification is:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{u}. \qquad (1)$$

In this study, the dependent variable ($\boldsymbol{y}$) represents the suicide mortality rate recorded from 2011 to 2021. The independent variables ($\mathbf{X}$) encompass $\mathbf{X1}$ to $\mathbf{X6}$. The estimated regression coefficients for these variables take the notation $\boldsymbol{\beta}$, and the error term takes the notation $\mathbf{u}$.

If spatial dependence exists in the underlying data, the OLS regression coefficients will be biased and/or the error term will be enlarged; in either case, this results in an imprecise estimation of the underlying relationships between the variables.

Spatial dependence can be explicitly modeled in a variety of ways. The spatial autoregressive (SAR) or spatial lag model inserts a parameter that captures spatial autocorrelation in the dependent variable, that is,

$$\mathbf{y} = \boldsymbol{\rho}\mathbf{W}\mathbf{y} + \mathbf{X\beta} + \mathbf{u}, \qquad (2)$$

where $\mathbf{W}$ is a spatial weights matrix that captures the spatial neighborhood of each observation. The weight matrix represents the spatial relationships between 31 observations (provinces) over the course of 11 years, considering the proximity of each observation to the other 30 provinces. Additionally, the spatial autoregressive parameter, represented by $\boldsymbol{\rho}$, captures the influence of spatial spillover effects on the dependent variable. By incorporating $\boldsymbol{\rho}$ in the model, we can analyze how the attributes of the neighboring provinces influence the dependent variable's values and gain insights into the spatial patterns and dynamics present in the dataset.

Another approach to modeling spatial dependence is to not to incorporate spatially lagged covariates in the equation. The spatial lag specification (SLX) for $\mathbf{X}$ is defined as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{WX\theta} + \mathbf{u}. \qquad (3)$$

In this model, $\boldsymbol{\theta}$ represents a vector of spatial spillover parameters. Beyond considering the direct effects of covariates, the model considers the additional influence from neighboring units' covariates, capturing the indirect spillover effects.

A noteworthy aspect of the model is its incorporation of spatial effects for each covariate, which the $\boldsymbol{\theta}$ vector encompasses. This feature enables the exploration of the spatial relationships and dependencies among the covariates, providing a comprehensive understanding of how neighboring units' characteristics contribute to the overall model dynamics.

In contrast, the Spatial Durbin Model (SDM) combines the spatial spillover specification of the covariates with the spatial autoregressive term of the dependent variable, yielding:

$$\mathbf{y} = \boldsymbol{\rho}\mathbf{W}\mathbf{y} + \mathbf{X\beta} + \mathbf{WX\theta} + \mathbf{u}. \qquad (4)$$

In practice, selecting the most appropriate model specification can be difficult, particularly when strong theoretical justifications are absent (25, 26). The primary objective of this paper is to assess the performance of traditional model specifications compared to random forest regressors. We evaluate the models based on prediction accuracy under two distinct testing scenarios and perform the estimation using R version 4.2.2.

Random Forests: Random forests, an ensemble learning technique, stand as a formidable tool in predictive modeling, amalgamating insights from multiple decision trees to refine predictions. At its core lies the Classification and Regression Tree (CART) training algorithm, which orchestrates the intricate dance of data partitioning and criterion optimization (12).

Decision trees, the elemental units of random forests, wield the power to address classification and regression tasks. In this study, we primarily focus on predicting nominal outcomes classified as 'Low', 'Medium', and 'High'. This focus naturally directs our analysis toward classification tree methods.

CART operates as a nonlinear function, sculpting the data landscape through a series of splits aimed to minimize impurity measures such as the Gini index or entropy (27). These criteria are commonly used in classification trees to maximize the purity of the resulting subsets after each split, which is crucial for accurate classification tasks.

Random forest harnesses the wisdom of crowds, leveraging the law of large numbers to refine predictions. By assembling an ensemble of decision trees, typically around 1,000, trained on random subsets of the data with replacement, random forest mitigate overfitting and enhance generalization capacity. The amalgamation of predictions from these individual trees through averaging yields a robust final prediction (28).

To promote diversity and independence among individual predictors, random forests restrict each tree's access to explanatory variables. By considering only a random subset of variables at each split, the classifier encourages diversity among predictors, which enhances the ensemble's overall predictive performance (2, 12).

Random forest's versatility extends beyond its robustness to noisy data and resilience to outliers. It offers interpretability through feature importance rankings, allowing users to understand the contribution of each variable to the prediction process. Furthermore, its scalability makes it suitable for large datasets and

parallel processing environments, ensuring efficient computation (28).

In our analysis, we embarked on a journey through four distinct random forest (RF) specifications to unravel the nuances of spatial factors in predictive modeling. The baseline model ("RF") mirrored the covariates utilized in the preceding spatial econometric models. Furthermore, we introduced the "RFSAR" model, which was engineered to emulate the spatial autoregressive (SAR) model by integrating the spatially lagged dependent variable. The "RFSDM" model extended its reach to encompass spatial lags of the dependent and independent variables, striving to mirror the spatial Durbin model (SDM). Lastly, the "RFSLX" model focused solely on the spatial lags of independent variables, akin to the spatial lag of the **X** model. By scrutinizing RF models through various spatial lenses, we unravel valuable insights into the intricate dance between spatial dynamics and predictive prowess, enriching our understanding and honing our predictive capabilities (15, 29, 30).

It is essential to clarify the distinction between incorporating spatiality in our random forest models and traditional spatial models. In our approach, spatially lagged variables were engineered as features and included in the RF training process to provide spatial context. This approach enables the model to indirectly capture spatial spillover effects, using the flexibility of machine learning to identify nonlinear relationships between neighboring provinces and suicide mortality rates. However, unlike conventional spatial models—such as spatial regression or spatial survival models—spatial dependence in RF is not intrinsic to the model's theoretical formulation or computational mechanics (31). In traditional spatial models, the likelihood or covariance structure explicitly incorporates spatial relationships, modeling spatial autocorrelation as an integral part of the estimation process. In contrast, our feature-engineered approach relies on the predictive power of RF to learn from spatially informative inputs, providing a practical but conceptually distinct mechanism for incorporating spatiality into predictive modeling. This distinction highlights the complementary nature of our hybrid framework: it combines the flexibility and nonlinear modeling capabilities of random forests with explicit spatial information, without requiring a fully parametric spatial dependence structure.

Overall, when constructing the hybrid random forest models, we incorporate spatial dependence directly at the data level. Specifically, spatially lagged variables were generated using the queen-contiguity weight matrix and added as input features to the dataset before model training (18). We did not modify the random forest algorithm, and we kept the tree-splitting criteria, such as the Gini index, unchanged (3). As a result, spatial spillover effects are captured indirectly through feature selection, rather than being explicitly included in the objective function. This approach maintains the robustness and reproducibility of the original random Forest framework, while enabling the model to incorporate spatial information effectively. Researchers have successfully applied similar strategies to enrich random forest models with spatially derived features in geoscience, environmental, and climate prediction tasks (15, 16, 30, 32), which further supports the validity of this method.

Results

In this study, we applied three spatial econometric models—SAR, SLX, and SDM—together with four data mining models—RF, RFSAR, RFSLX, and RFSDM—to classify provinces based on suicide mortality rates. These models classify outcomes into three categories: Low, Medium, and High. We applied these models to predict suicide mortality levels, categorizing each observation as Low, Medium, or High. This rigorous approach ensured a thorough assessment of the models' classification ability.

Table 3 provides a comprehensive evaluation of the models' performance across various metrics, facilitating comparison and identification of superior models in terms of accuracy, precision, sensitivity, F-score, and specificity. Notably, RFSDM emerged as the standout model based on these criteria.

Table 3. Performance Metrics Calculated on the Entire Dataset: A Comparative Analysis of Prediction Models for Suicide Mortality Levels

Model	Accuracy	Precision	Sensitivity	F-score	Specificity
SAR	0.3659	0.3750	0.2727	0.3158	0.4737
SLX	0.4651	0.4500	0.4286	0.4390	0.5000
SDM	0.4651	0.4211	0.4000	0.4103	0.5217
RF	0.2703	0.0769	0.4000	0.1290	0.2500
RFSAR	0.8250	0.4615	1	0.6316	0.7941
RFSLX	0.8684	0.4444	1	0.6154	0.8529
RFSDM	0.8919	0.4286	1	0.6000	0.8824

Accuracy measures the overall correctness of the model's predictions, calculated as the ratio of correctly predicted instances to the total number of cases. For example, the RFSLX model achieved an accuracy of 0.8684, indicating it correctly predicted 86.84% of cases.

Precision assesses the proportion of accurate optimistic predictions out of all optimistic predictions made by the model. Higher precision values signify fewer false positives. The RFSDM model demonstrated a precision of 0.4286, indicating 42.86% of predicted positive instances were true positives.

Sensitivity, also known as recall or actual accurate rate, measures the proportion of accurate optimistic predictions from all actual positive instances, indicating how well the model identifies positive instances. The RFSAR model displayed a sensitivity of 1, accurately identifying all positive instances.

The F-score, the harmonic mean of precision and sensitivity, evaluates the balance between these metrics. A higher F-score indicates better performance in handling false positives and false negatives. For instance, the RFSAR model achieved an F-score of 0.6316, indicating balanced performance.

Specificity measures the proportion of accurate pessimistic predictions from all actual negative instances, indicating how well the model identifies negative instances. Higher specificity values suggest fewer false positives. For example, the RFSDM model demonstrated specificity values of 0.8824, correctly identifying 88.24% of negative cases.

In summary, these performance metrics, calculated on the entire dataset, provide valuable insights into the predictive capabilities of each model, aiding in informed decision-making and model selection for predicting suicide mortality levels.

The confusion matrix and the out-of-bag (OOB) error plot are pivotal evaluation tools for assessing the performance of a random forest model in classification tasks. While the confusion matrix provides a detailed breakdown of the model's predictions for each class, aiding in assessing accuracy, precision, sensitivity, specificity, and other metrics, the OOB error plot visualizes the OOB error rate. This rate, estimating the model's prediction error on unseen data, is plotted against the number of trees in the random forests. This allows us to gauge the model's overall performance and identify the optimal number of trees. In Figure 1, the x-axis represents the number of trees, and the y-axis represents the OOB error rate, which exhibits a notable decrease with an increasing number of trees.

However, the OOB error rate eventually stabilizes, around 0.09 or 9% after approximately 30 trees, indicating diminishing returns beyond this point. Thus,

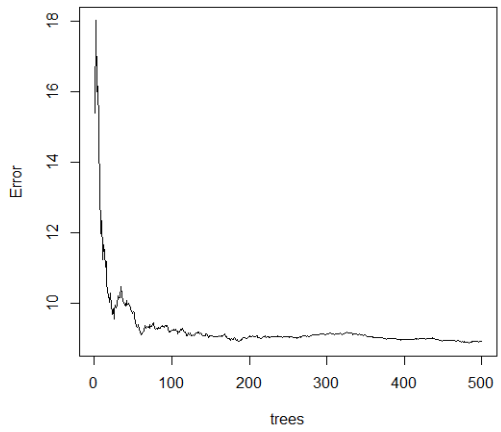

Figure 1. Out-of-Bag (OOB) Error Rate Trend for RFSDM.

the OOB error plot assists in determining the optimal number of trees for our random forest model (RFSDM), striking a balance between capturing patterns and avoiding overfitting.

By selecting an appropriate number of trees, typically around 30 in our case, we ensure that our RFSDM model performs well on unseen data while minimizing computational complexity. In conclusion, analyzing the OOB error rate trend provides useful insights into model performance and supports more informed research decisions.

Table 4 displays the confusion matrices for the six models utilized in this study to forecast suicide mortality levels categorized as Low, Medium, and High. The evaluation of each model's performance compares its predicted outcomes with the actual observations.

The table delineates the instances classified into each level for the corresponding model's predictions, along with the total counts of correct and incorrect predictions made by each model. The analysis examines the SAR, SLX, SDM, RF, RFSAR, RFSLX, and RFSDM models and carefully documents their correct and incorrect predictions. These confusion matrices provide invaluable insights into the efficacy of each model, facilitating an assessment of their accuracy and error rates.

Upon comparison of these models' performance, it is evident that RFSDM surpasses the others in predicting suicide mortality levels, demonstrating superior accuracy and overall predictive performance. Specifically, the model correctly predicts 92 instances: 30 in the Low category, 29 in the Medium category, and 33 in the High category. However, it erroneously predicts 12 cases across all levels.

Table 5 illustrates the cross-tabulation of the RFSDM Model's predictions with the actual suicide mortality levels within the dataset. The table cells contain various

Table 4. Confusion Matrix using the Seven Models

Model	Level	Low	Medium	High	Correct	Incorrect
SAR	Low	9	6	13	33	71
	Medium	10	15	12		
	High	16	14	9		
SLX	Low	11	9	9	36	68
	Medium	11	11	12		
	High	12	15	14		
SDM	Low	12	8	9	37	67
	Medium	11	11	12		
	High	12	15	14		
RF	Low	8	2	0	56	48
	Medium	24	25	12		
	High	3	8	23		
RFSAR	Low	27	6	0	86	18
	Medium	7	26	2		
	High	0	3	33		
RFSLX	Low	29	4	0	90	14
	Medium	5	28	2		
	High	0	3	33		
RFSDM	Low	30	3	0	92	12
	Medium	4	29	2		
	High	0	3	33		

Table 5. Confusion Matrix: Predicted vs. Suicide Mortality Levels

Categories	Low	Medium	High	Row Total
	30	3	0	
	34.211	5.916	11.106	33
Low	0.909	0.091	0.000	
	0.882	0.086	0.000	0.317
	0.288	0.029	0.000	
	4	29	2	35
	4.841	25.178	8.118	
Medium	0.114	0.829	0.057	0.337
	0.118	0.829	0.057	
	0.038	0.279	0.019	
	0	3	33	
	11.769	6.858	36.001	36
High	0.000	0.083	0.917	
	0.000	0.086	0.943	0.346
	0.000	0.029	0.317	
Column Total	34	35	35	104
	0.327	0.337	0.337	

information, including the number of observations (N), the Chi-square contribution, the proportion of observations relative to the row total (N / Row Total), the proportion of observations relative to the column total (N / Col Total), and the proportion of observations relative to the overall total (N / Table Total).

The cross table encompasses a total of 104 observations from the dataset. It is important to emphasize that these 104 instances constitute a comprehensive evaluation of the model's performance across all categories and provinces, ensuring a thorough assessment of the predictive accuracy and effectiveness of the random forest models in predicting suicide mortality levels in the Iranian provinces. The cross table is structured with rows representing the predicted suicide mortality levels by the model and columns denoting the actual suicide mortality levels. Additionally, the rightmost column showcases the row totals, indicating the total number of observations for each predicted suicide mortality level. Furthermore, the bottom row provides the column totals, representing the cumulative number of observations for each suicide mortality level.

To interpret the table, we scrutinize the values within each cell. For example, in the first row, the model predicts 30 instances as Low, three as Medium, and none as High. The row total, up to 33, signifies the total

number of predictions made for the Low category.

The Chi-square contribution values in Table 5 reflect each cell's impact on the overall Chi-square statistic, a metric assessing the model's goodness of fit. This measure evaluates how well the model fits the data when predicting suicide mortality levels across the categories of Low, Medium, and High.

These values indicate how much each cell contributes to the overall Chi-square statistic, which measures the discrepancy between the observed and expected frequencies in the data.

Similarly, we can interpret the other rows and columns to gauge the model's performance in predicting suicide mortality levels. Overall, the table furnishes invaluable insights into the model's accuracy and predictive prowess, revealing the degree of alignment between the model's predictions and the actual data.

Conclusion

In this study, we advanced the predictive modeling of suicide mortality across Iranian provinces by comparing traditional spatial econometric models (SAR, SLX, and SDM) with random forest–based methods. We also introduced a hybrid framework that integrates spatially lagged features into random forests. This framework combines the strengths of both approaches: the predictive accuracy of random forests in capturing complex nonlinear interactions among covariates (2, 12) and the interpretability of spatial econometric models in uncovering structured spatial dependencies and spillover effects (21, 33). The strong performance of the RFSDM model shows that embedding spatial information into machine learning improves prediction accuracy while preserving interpretability that supports evidence-based decision-making (15, 32).

Our findings extend the literature on suicide prediction by demonstrating the value of spatially informed machine learning. While earlier studies focused mainly on socio-demographic and environmental correlates of suicide (34, 35, 36), our hybrid approach uncovers additional spatial patterns and risk structures that remain hidden in purely econometric models (22) or non-spatial machine learning models (28). These findings underscore the importance of incorporating spatial context into predictive modeling, especially when neighboring regions influence outcomes.

The practical implications for public health policy are substantial. By accurately identifying high-risk provinces and anticipating spatial spillovers in suicide risk, policymakers can allocate mental health resources more efficiently and design targeted, region-specific prevention strategies. The hybrid framework serves as a decision-support tool by predicting suicide risk and guiding intervention planning, thereby enhancing the effectiveness of evidence-based prevention programs in Iran and potentially in other regions (World Health Organization, 2021; 37).

Despite these contributions, the study has several limitations. Aggregating data at the provincial level may obscure intra-provincial variability, and the choice and availability of predictor variables may influence model performance. Additionally, temporal dynamics were not explicitly incorporated in the current analysis. Future research should explore richer covariates, test alternative spatial machine learning algorithms, integrate dynamic spatial models, and extend the framework to other health outcomes (7, 30). Addressing these avenues will refine spatially explicit predictive modeling and enhance its applicability for public health decision-making.

In summary, this study makes three key contributions: (i) development of a hybrid RF–spatial econometric framework that balances predictive accuracy and interpretability; (ii) demonstration that incorporating spatial lags into machine learning substantially improves suicide risk prediction; and (iii) provision of actionable insights for public health policy, enabling targeted, data-driven suicide prevention strategies. These contributions highlight the value of combining traditional spatial modeling with modern machine learning approaches for health outcomes research.

Funding

No funding sources supported our research.

Conflicts of Interest

We declare no potential conflicts of interest or competing interests related to the research.

References

1. Openshaw S, Openshaw C. Artificial intelligence in geography. New York (NY): John Wiley & Sons Inc; 1997.
2. Breiman L. Random forests. Mach Learn. 2001;45:5-32.
3. Breiman L. Statistical modeling: the two cultures. Stat Sci. 2001;16:199-231.
4. Shojaei A, Moradi S, Alaeddini F, Khodadoost M, Barzegar A, Khademi A. Association between suicide method and gender, age, and education level in Iran over 2006–2010. Asia Pac Psychiatry. 2014;6:18-22.
5. Khademi AMS. Statistical investigation of unnatural deaths in Iran (2001–2010). Tehran: Forensic Medicine Organization; 2012.
6. Rey S. Geographical analysis: reflections of a recovering editor. Geogr Anal. 2019;1-9. doi:10.1111/gean.12193.
7. Singleton A, Arribas-Bel D. Geographic data science. Geogr Anal. 2019;1-15. doi:10.1111/gean.12194.

8. Adin A, Retegui G, Villegas AS, Ugarte MD. Suicide Mortality in Spain (2010-2022): Temporal Trends, Spatial Patterns, and Risk Factors. arXiv preprint arXiv:2509.01342. 2025.
9. Al-Remawi M, Agha ASA, Al-Akayleh F, Aburub F, Abdel-Rahem RA. Artificial intelligence and machine learning techniques for suicide prediction: Integrating dietary patterns and environmental contaminants. Heliyon. 2024;10(24).
10. Ehtemam H, Sadeghi Esfahlani S, Sanaei A, Ghaemi MM, Hajesmaeel-Gohari S, Rahimisadegh R, et al. Role of machine learning algorithms in suicide risk prediction: a systematic review and meta-analysis of clinical studies. BMC Med Inform Decis Mak. 2024;24(1):138.
11. Prateek J. Artificial intelligence with Python: a comprehensive guide to building intelligent apps for Python beginners and developers. Birmingham (UK): Packt Publishing; 2017. p. 49-67.
12. Géron A. Hands-on machine learning with Scikit-Learn & TensorFlow: concepts, tools, and techniques to build intelligent systems. Boston (MA): O'Reilly Media; 2017. p. 3-224.
13. Janowicz K, Gao S, McKenzie G, Hu Y, Bhaduri B. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. Int J Geogr Inf Sci. 2019;34:625-36. doi:10.1080/13658816.2019.1684500.
14. Credit K. Transitive properties: a spatial econometric analysis of new business creation around transit. Spatial Econ Anal. 2019;14:26-52.
15. Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B. Random forests as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ. 2018;6:e5518. doi:10.7717/peerj.5518.
16. Georganos S, Grippa T, Gadiaga N, Linard C, Lennert M, Vanhuysse S, et al. Geographical random forests: a spatial extension to the random forests algorithm to address spatial heterogeneity in remote sensing and population modeling. Geocarto Int. 2019. doi:10.1080/10106049.2019.1595177.
17. Yan B, Janowicz K, Mai G, Zhu R. A spatially explicit reinforcement learning model for geographic knowledge graph summarization. Trans GIS. 2019;29:620-40.
18. Farzammehr MA, Moradi S. Classical and spatial cluster analysis of smuggling in Iranian provinces. Reg Sci Policy Pract. 2022;16(2):12609.
19. Anselin L. Spatial econometrics: methods and models. Dordrecht: Kluwer Academic Publishers; 1988. p. 169-79.
20. Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: data mining and knowledge discovery. 2019 May;9(3):e1301.
21. Elhorst JP. Spatial econometrics: from cross-sectional data to spatial panels. Heidelberg: Springer; 2014 Jan 16.
22. LeSage JP, Pace RK. Introduction to spatial econometrics. New York (NY): CRC Press; 2009. p. 25-75.
23. Haining RP. Spatial data analysis: theory and practice. Cambridge university press; 2003 Apr 17.
24. Tobler WR. A computer movie simulating urban growth in the Detroit region. Econ Geogr. 1970;46:234-40.
25. Anselin L, Rey S. Modern spatial econometrics in practice. Chicago (IL): GeoDa Press LLC; 2014. p. 1-43.
26. LeSage JP. Spatial econometric panel data model specification: a Bayesian approach. Spatial Stat. 2014;9:122-45.
27. Krzywinski M, Altman N. Classification and regression trees. Nat Methods. 2017;14:757-8.
28. Fox EW, Ver Hoef JM, Ohlen AR. Comparing spatial regression to random forests for large environmental data sets. PLoS One. 2020;15(3):e0229509.
29. Lagomarsino D, Tofani V, Segoni S, Catani F, Casagli N. A tool for classification and regression using random forest methodology: applications to landslide susceptibility mapping and soil thickness modeling. Environ Model Assess. 2017;22:201-14.
30. Talebi H, Peeters LJ, Otto A, Tolosana-Delgado R. A truly spatial random forests algorithm for geoscience data analysis and modelling. Math Geosci. 2022;54(1):1-22.
31. Madadi M, Motarjem K. Survival data analysis using statistical learning methods. J Stat Sci. 2025;18(2).
32. Wang Y, Yin Y, Gao B, Zeng Y, Zhao Y, Chen Z, et al. A novel spatial prediction method integrating Exploratory Spatial Data Analysis into Random Forest for large-scale daily air temperature mapping. IEEE Trans Geosci Remote Sens. 2025.
33. Anselin L, Bera AK, Florax R, Yoon MJ. Simple diagnostic tests for spatial dependence. Reg Sci Urban Econ. 1996;26:77-104. doi:10.1016/0166-0462(95)02111-6.
34. Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S. Suicide and suicidal behavior. Epidemiologic reviews. 2008 Jul 24;30(1):133.
35. Bilsen J. Suicide and youth: risk factors. Frontiers in psychiatry. 2018 Oct 30;9:540.
36. Credit K. Spatial models or random forests? Evaluating spatially explicit machine learning methods to predict employment density around new transit stations in Los Angeles. Geogr Anal. 2022;54:58-83.
37. Stack S. Contributing factors to suicide: Political, social, cultural and economic. Preventive medicine. 2021 Nov 1;152:106498.

Journal of Sciences, Islamic Republic of Iran 36(1): 51 - 59 (2025) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Precision Tuning of a kHz-Driven Argon Plasma Jet Enables Dose-Controlled H_2O_2 Delivery to Overcome Chemoresistance in Colorectal Cancer

A. Eftekharinasab*, H. Mehdian, A. Hasanbeigi

Department of Physics and Institute for Plasma Research, Kharazmi University, Tehran, Islamic Republic of Iran

Received: 3 September 2025 / Revised: 20 October 2025 / Accepted: 28 October 2025

Abstract

Colorectal cancer presents a significant therapeutic challenge, largely due to robust chemoresistance mechanisms, including the upregulation of antioxidant pathways. While cold atmospheric plasma is a promising anti-cancer modality, its efficacy can be limited by these cellular defenses. This study introduces a kilohertz AC-driven argon plasma jet with independently tunable voltage (1–20 kV) and frequency (18–28 kHz) as a novel platform for overcoming this resistance. We demonstrate that precision tuning of these electrical parameters allows for the controlled delivery of extracellular hydrogen peroxide (H_2O_2), a key long-lived reactive species. In the chemoresistant HT29 colorectal cancer cell line, we achieved a modulation of H_2O_2 concentrations in the culture medium, ranging from 291 to 371 µM. This H_2O_2 dosage showed a linear correlation with dose-dependent cytotoxicity ($R^2 = 0.995$, $p < 0.001$). Optimized parameters (10.5 kV, 28 kHz) overwhelmed the cells' redox defenses, reducing viability to 9.2% ± 3.6% after a 3-minute treatment. This approach successfully bypasses the Nrf2/Srx antioxidant pathway, which is known to confer resistance to helium plasma jets. Our findings establish that precisely controlling H_2O_2 delivery via a tunable argon plasma jet is a potent strategy for circumventing intrinsic chemoresistance in colorectal cancer, positioning this technology as a promising modality for precision oncology.

Keywords: Argon plasma jet; Colorectal cancer; Chemoresistance; Hydrogen peroxide (H_2O_2).

Introduction

Colorectal cancer (CRC) remains a leading cause of cancer-related mortality worldwide, with projections indicating a growing global burden (1, 2). Clinical outcomes are often compromised by high recurrence rates, driven by intrinsic therapeutic resistance and the molecular heterogeneity of tumors (3, 4). This persistent challenge underscores the urgent need for innovative therapeutic strategies capable of overcoming cellular resistance mechanisms to improve patient outcomes (5).

Cold atmospheric plasma (CAP), a partially ionized gas generated at near-body temperature, has emerged as a promising modality in oncology (6-9). CAP produces a complex cocktail of reactive oxygen and nitrogen species (RONS), including both short-lived radicals (e.g., •OH,

* Corresponding Author: Tel: +982186072780; Email: eftekharinasab@gmail.com

$^1O_2$) and long-lived molecules (e.g., H_2O_2, NO_2^-) (10). The therapeutic principle relies on inducing overwhelming oxidative stress in cancer cells, which often exhibit a dysregulated redox homeostasis, thereby triggering apoptosis (11). Among these species, hydrogen peroxide (H_2O_2) is recognized as a key long-lived effector molecule in CAP-mediated anticancer activity, capable of diffusing through tissues and initiating sustained cytotoxic signaling (12).

The biological effect of CAP is not uniform; it is critically dependent on the device configuration (e.g., dielectric barrier discharge (DBD) vs. plasma jet) and operating parameters, which dictate the composition and flux of the generated RONS (13-16). However, a significant hurdle for CAP therapy is the adaptive response of chemoresistant cancer cells. The HT29 CRC cell line, for example, is notoriously resistant to oxidative stress due to its upregulation of potent antioxidant defense pathways, particularly the Nrf2/Sulfiredoxin (Srx) axis (17). A seminal study by Ishaq et al. (2014) demonstrated that this robust antioxidant system allowed HT29 cells to effectively neutralize RONS generated by a helium plasma jet, thus mitigating its therapeutic effect (17). This finding underscores the critical need for a CAP system capable of delivering a RONS dosage potent and controlled enough to overwhelm these protective cellular mechanisms.

To contextualize our approach, it is crucial to review prior studies investigating CAP's efficacy on the HT29 cell line, which are summarized in Table 1. The initial pivotal study by Ishaq et al. utilized a helium-based plasma jet and revealed that HT29 cells exhibit significant resistance due to the upregulation of the Nrf2/Srx antioxidant pathway (17). This resistance was only overcome when the protective pathway was silenced via siRNA, highlighting the need for a more potent CAP source. Subsequent research predominantly employed DBDs operating with ambient air. These studies, often conducted by researchers in the biological sciences, successfully induced apoptosis in HT29 cells without genetic intervention by exploring various molecular pathways, including p53-independent mechanisms and cell cycle arrest (18-21). For instance, studies utilized commercial devices like the Piezobrush® PZ2 and miniFlatPlaSter, which were originally designed for industrial surface processing or other medical applications and operate with fixed parameters (19, 20). Most recently, Martinet et al. investigated a commercial argon-based electrosurgical device (APC 3 VIO 3), originally intended for hemostasis (blood coagulation), and demonstrated its cytotoxic potential in a "cold" operating mode (22). While these studies confirmed the general anti-cancer potential of CAP, they largely relied on devices with fixed or limitedly adjustable operating parameters. This "one-size-fits-all" approach does not account for the specific redox threshold of highly resistant cells like HT29. This underscores a critical limitation: the lack of a systematic approach to tune plasma chemistry to specifically overwhelm known resistance mechanisms.

Our work directly addresses this limitation by introducing a fully tunable kHz AC-driven argon plasma jet engineered for this precise purpose. We hypothesized that the precise control afforded by this system would enable a targeted and dose-controlled delivery of H_2O_2 sufficient to surpass the antioxidant threshold of resistant HT29 cells. Unlike fixed-parameter systems (such as radio-frequency (RF) jets) or less targetable DBDs, our approach allows for the systematic optimization of plasma chemistry to maximize cytotoxicity. This study

Table 1. Summary of previous studies investigating the effects of CAP on the HT29 colorectal cancer cell line.

Study (year)	CAP Source/ Device	Working Gas	Key Findings	Ref.
Ishaq (2014)	Plasma jet	Helium	HT29 cells are resistant via the Nrf2/Srx pathway; cytotoxicity requires siRNA silencing of this pathway.	(17)
Han (2017)	DBD	Air	Induces cell cycle arrest and apoptosis via regulation of Sp1 transcription factor.	(21)
Schneider (2018)	DBD/miniFlatPlaSter	Air	Inhibits cancer cell growth regardless of p53 mutation status; non-toxic to normal colon tissue	(20)
Wang (2022)	DBD/Piezobrush® PZ2	Air	Induces mitochondrial apoptosis pathway (cytochrome c, caspase-9/3 activation).	(19)
He (2024)	DBD	Air	Induces apoptosis via the cytochrome c-caspase-9-caspase-3 pathway, independent of microsatellite instability status.	(18)
Martinet (2025)	Electrosurgical Cold Plasma Device/APC 3 VIO 3	Argon	"Cold mode" operation generates ROS (especially H_2O_2) and induces dose-dependent cytotoxicity.	(22)

aims to: (1) characterize the voltage- and frequency-dependent generation of H_2O_2 by our tunable argon plasma jet; (2) quantify the resulting cytotoxicity in the chemoresistant HT29 cell line; and (3) establish a predictive model linking the H_2O_2 dose to cell death, thereby demonstrating a novel strategy to bypass a well-defined molecular resistance mechanism in CRC.

Materials and Methods

Plasma Jet Design and Power Supply

A custom-fabricated argon plasma jet was developed to generate a spatially uniform and stable nonthermal plasma plume for RONS delivery. The jet body, constructed from acrylonitrile butadiene styrene (ABS), housed a quartz dielectric tube (3 mm inner diameter, 5 mm outer diameter). A stainless-steel rod (1.8 mm diameter), serving as the high-voltage electrode, was inserted into the quartz tube. A 5 mm-long grounded copper ring electrode was positioned on the tube's exterior, such that its distal edge was 4 mm from the tube exit (Figure 1). This coaxial electrode configuration, with the central electrode aligned within the axial plane of the ring, ensured a symmetrical and consistent plasma discharge. The system operated with 99.999% pure argon, regulated at 2 standard liters per minute (SLM) using a mass flow controller. Plasma excitation was achieved using a custom-built kHz AC power supply (1–20 kV, 18–28 kHz) with independent voltage and frequency tuning. Electrical diagnostics identified a resonant frequency range of 20–24 kHz, corresponding to maximum energy transfer to the plasma column.

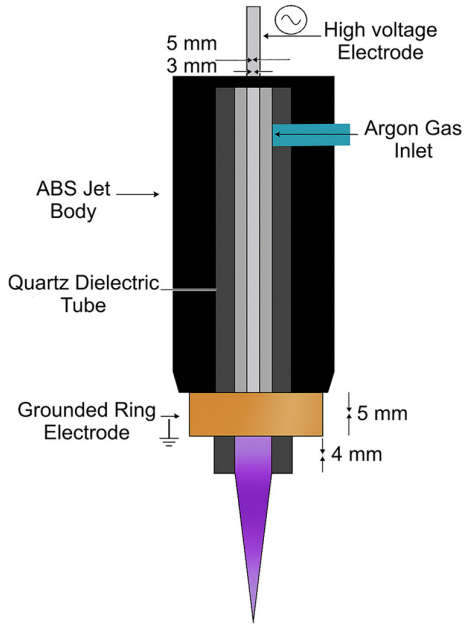

Figure 1. Schematic representation of the Argon Plasma Jet nozzle.

Cell Culture

The human colorectal adenocarcinoma cell line HT29 (IBRC, C10097) was cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. The cells were maintained in a humidified incubator at 37°C with 5% CO_2. For experiments, cells in the logarithmic growth phase were detached using 0.25% trypsin-EDTA, and seeded in 35 mm culture dishes at a density of 2×10^5 cells/dish in 2 mL of complete medium. After 24 hours of incubation, cultures reached 70–80% confluency.

Plasma Treatment

Prior to treatment, HT29 cells cultured in DMEM with 10% heat-inactivated FBS were washed twice with Phosphate-buffered saline (PBS). For viability assays, cells were incubated in 2 mL of serum-free DMEM to avoid FBS-mediated scavenging of plasma-generated reactive species. For H_2O_2 quantification, 2 mL of DMEM (without FBS) was treated under identical conditions in cell-free dishes. The plasma jet nozzle was positioned 10.0 mm vertically above the liquid surface (Figure 2). Cells were exposed to argon plasma at seven distinct voltage-frequency combinations during the screening phase, including (9.5 kV, 18 kHz), (10.5 kV, 18 kHz), (9.5 kV, 25.5 kHz), (10.5 kV, 25.5 kHz), (11.1 kV, 25.5 kHz), (9.5 kV, 28 kHz), and (10.5 kV, 28 kHz) , each applied for 3 minutes. For time-dependency analysis, the optimized parameters of (10.5 kV, 28 kHz) were applied for 1 to 4 minutes. Argon (99.999% purity) continuously supplied at 2 SLM. Four hours after plasma exposure,

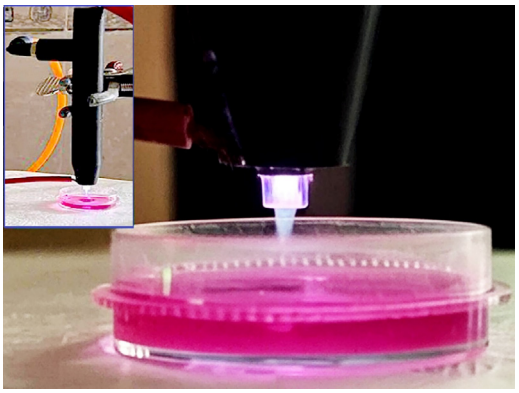

Figure 2. Argon plasma jet treatment in a 35 mm culture dish.

0.22 mL of heat-inactivated FBS was added to restore 10% serum concentration, standardizing post-treatment conditions for viability assays. All experiments were conducted in triplicate biological replicates, with untreated serum-free controls processed in parallel.

Cell Viability Assay (MTT)

Cell viability was assessed 24 hours post-treatment using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay. After plasma exposure, cells were incubated with 0.5 mg/mL MTT in PBS for 4 hours at 37°C under 5% CO_2. The supernatant was gently removed, and the resulting formazan crystals were solubilized in 1 mL dimethyl sulfoxide (DMSO). Absorbance at 570 nm was recorded using a spectrophotometric microplate reader. All measurements were performed in triplicate.

H_2O_2 Concentration Measurement

H_2O_2 concentrations in cell-free DMEM (2 mL) were quantified using a colorimetric hydrogen peroxide assay kit (ZellBio GmbH) following the manufacturer's protocol. All plasma treatment conditions were replicated identically in the quantification assays.

Statistical Analysis

Statistical analyses were performed using Origin 10.

Results

Voltage- and Frequency-Dependent Cytotoxicity

Systematic modulation of voltage (9.5–11.1 kV) and frequency (18–28 kHz) revealed distinct cytotoxic profiles in HT29 cells following 3- minute exposure to the argon plasma jet (Figure 3). At 18 kHz, increasing the voltage from 9.5 kV to 10.5 kV reduced viability from 25.16% ± 3.9% to 19.87% ± 3.2%, demonstrating voltage-dependent enhancement in cytotoxicity. In contrast, at 25.5 kHz, all tested voltages (9.5, 10.5, 11.1 kV) yielded comparatively higher viability rates (viability: 35–48%), suggesting frequency-mediated attenuation of reactive species generation. Notably, Maximum cytotoxicity (9.2% ± 3.6% viability) was observed at 10.5 kV and 28 kHz, while under identical frequency, reducing the voltage to 9.5 Kv resulted in lower cytotoxicity (54.3% ± 1.2% viability). These findings underscore the critical interplay between voltage and frequency in modulating CAP-induced cell death.

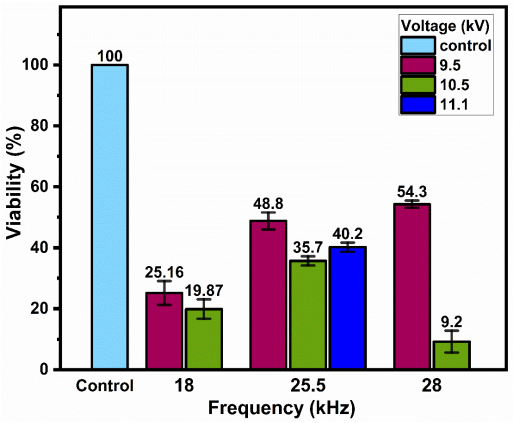

Figure 3. Voltage- and frequency-dependent cytotoxicity in HT29 cells after 3-minute plasma treatment. MTT assay results show cell viability (%) at different voltage-frequency combinations (9.5–11.1 kV, 18–28 kHz). Maximum cytotoxicity (9.2% viability) occurred at 10.5 kV, 28 kHz. Data: mean ± SD (n=3).

Time-Dependent Cytotoxicity at Optimal Parameters

Time-course experiments conducted at the optimized conditions (10.5 kV, 28 kHz) demonstrated rapid and saturable cytotoxic kinetics in HT29 cells (Figure 4a). Cell viability declined exponentially from 91.4% ± 3.9% after 1 minute of plasma exposure to 9.2% ± 3.6% after

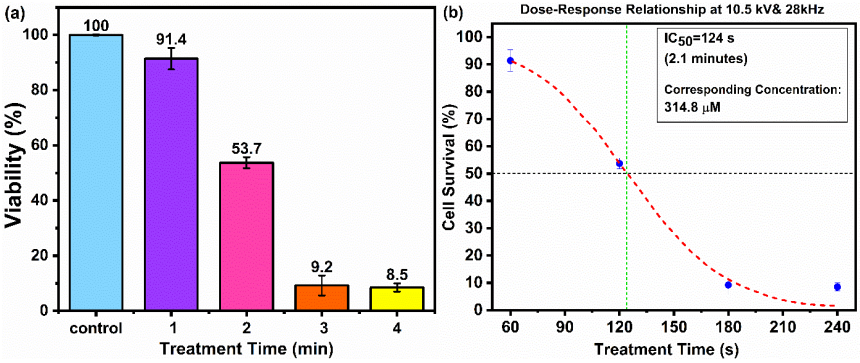

Figure 4. Time-dependent cytotoxicity at 10.5 kV, 28 kHz. **a**. Viability decreased progressively with exposure time (1–4 minutes). Data: mean ± SD (n=3). **b**. Correlation between H_2O_2 concentration and IC50 at 10.5 kV, 28 kHz. The IC50 treatment time (124 seconds) corresponds to 314.8 μM H_2O_2.

3 minutes, with an IC_{50} of 124 s (Figure 4b). Extending the exposure to 4 minutes resulted in 8.5% ± 1.5% viability, showing no statistically significant increase in cytotoxicity beyond the 3-minute mark. This plateau suggests that H_2O_2 accumulation may reach saturation, limiting additional cytotoxic effects despite prolonged treatment duration.

Tunable H_2O_2 Generation via Voltage-Frequency Modulation

Extracellular H_2O_2 concentrations measured in cell-free DMEM closely mirrored the observed cytotoxicity patterns. At 10.5 kV, H_2O_2 production reached its peak value of 371.0 ± 12.7 µM (at 28 kHz), surpassing the levels obtained at 25.5 kHz (319.15 ± 12.09 µM) and 18 kHz (347.75 ± 14.50 µM) (Figure 5). Frequency modulation produced distinct H_2O_2 generation profiles that were strongly dependent on the applied voltage (Figure 6). Specifically, at 10.5 kV, H_2O_2 concentrations followed a parabolic response, with a clear maximum at 28 kHz (Figure 7a). In contrast, at 9.5 kV, H_2O_2 output exhibited an exponential decay with increasing frequency, declining from 335.94 ± 16.59 µM (at 18 kHz) to 291.50 ± 9.19 µM (at 28 kHz) (Figure 7b). Notably, operation at 11.1 kV, 25.5 kHz yielded suboptimal H_2O_2 levels (310.55 ± 13.36 µM), possibly due to reactive species quenching under excessively energetic plasma conditions.

Temporal Kinetics of H_2O_2 Accumulation

Under optimized plasma conditions (10.5 kV, 28 kHz), H_2O_2 accumulation in cell-free DMEM exhibited time-dependent saturation behavior (Figure 8). The concentration increased progressively, reaching 371.0 ± 12.7 µM at 3 minutes, which corresponded closely with the observed IC_{50} value of 314 µM for HT29 viability. Extending plasma exposure to 5 minutes resulted in only a modest further increase in H_2O_2 levels (385.0 ± 16.5 µM), indicating a plateau phase.

H_2O_2 Drives Argon Plasma Jet-Induced Cytotoxicity

A strong, statistically significant linear correlation (R^2 = 0.995, $p < 0.001$) was observed between extracellular H_2O_2 concentrations (ranging from 291 to 371 µM) and HT29 cell viability across all tested voltage–frequency conditions (Figure 9). Regression analysis confirmed that H_2O_2 levels alone could reliably predict treatment-induced cytotoxicity, highlighting the potential for dose-controlled oxidative stress induction via precision tuning of the khz-driven argon plasma jet.

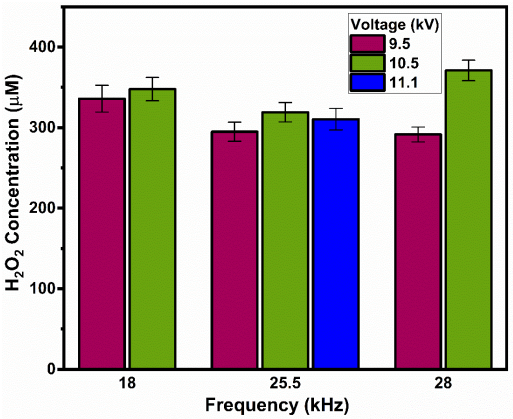

Figure 5. H_2O_2 generation across voltage-frequency combinations. H_2O_2 concentrations (µM) in cell-free DMEM after 3-minute exposure. Highest H_2O_2 (371.0 ± 12.7 µM) at 10.5 kV, 28 kHz. Data: mean ± SD (n=3).

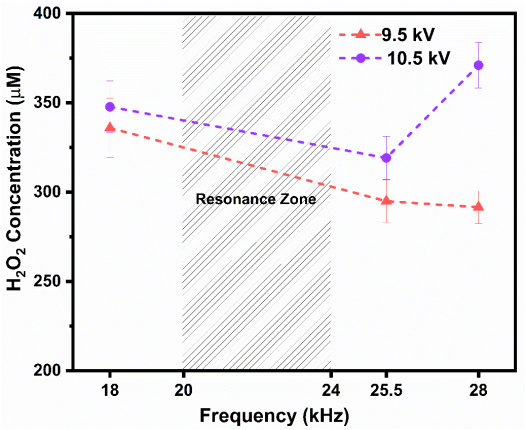

Figure 6. H_2O_2 concentration for 9.5 kV and 10.5 kV across frequency. At 10.5 kV, H_2O_2 generally exceeded 9.5 kV. Data: mean ± SD (n=3).

Discussion

This study demonstrates that a voltage- and frequency-tunable argon plasma jet can overcome chemoresistance in HT29 CRC cells through the precise, dose-controlled delivery of H_2O_2. Our principal finding is the exceptionally strong linear correlation (R^2 = 0.995, $p < 0.001$) between the extracellular H_2O_2 concentration and the reduction in cell viability. By optimizing the electrical parameters to 10.5 kV and 28 kHz, we generated a cytotoxic environment that reduced HT29 viability to just 9.2%, showcasing a level of efficacy not easily achieved with conventional, non-tunable plasma systems against this resistant cell line.

Overcoming the Antioxidant Defense

The significance of our findings is best understood in

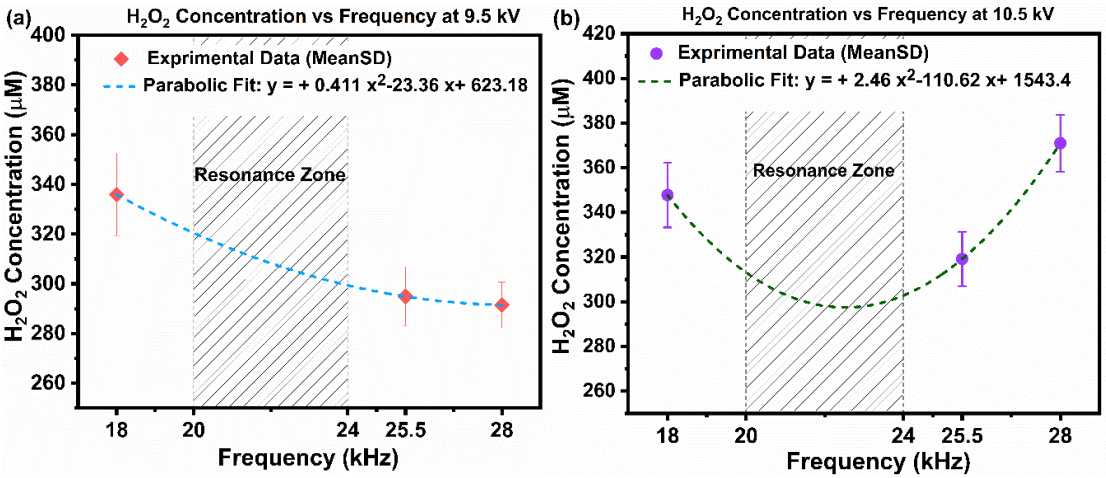

Figure 7. Frequency-dependent H_2O_2 trends. **a** At 9.5 kV, parabolic .**b** At 10.5 kV, parabolic. Data: mean ± SD (n=3).

the context of HT29's known resistance mechanisms. These cells exhibit a high basal expression of the Nrf2/Srx antioxidant axis, a protective system that effectively neutralizes ROS-induced damage and confers resistance to both chemotherapy and certain forms of CAP (17, 23-29). The work by Ishaq et al. (2014) was pivotal in showing that a helium plasma jet failed to induce significant apoptosis in HT29 cells unless the Nrf2/Srx pathway was silenced via siRNA (17). In stark contrast, our tunable argon plasma jet induces potent cytotoxicity without any genetic or chemical sensitization. We propose that by precisely tuning the voltage and frequency, we can generate a sufficiently high and sustained flux of H_2O_2 (up to 371 µM) that effectively saturates and overwhelms the buffering capacity of the Nrf2/Srx system. This ability to overcome a specific, well-documented resistance pathway represents a key novelty and a significant step forward for plasma cancer therapy.

The Interplay of Voltage and Frequency in Controlling Plasma Chemistry

The experimental data revealed a clear interplay between applied voltage and driving frequency in modulating H_2O_2 production within the argon plasma jet. Voltage dependence was evident, as higher voltages (e.g., 10.5 kV compared to 9.5 kV) promoted electron-impact reactions, leading to an increased •OH radical density ($Ar + e^- \rightarrow Ar^* + \bullet OH$) and subsequent H_2O_2 formation ($\bullet OH + \bullet OH \rightarrow H_2O_2$) (30). Our results indicate the existence

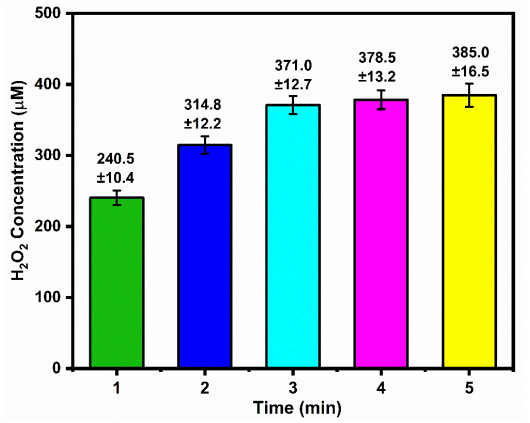

Figure 8. Time-dependent H_2O_2 accumulation at 10.5 kV, 28 kHz. Saturating increase in H_2O_2 (1–5 minutes). Plateau was observed after 3 minutes. Data: mean ± SD (n=3).

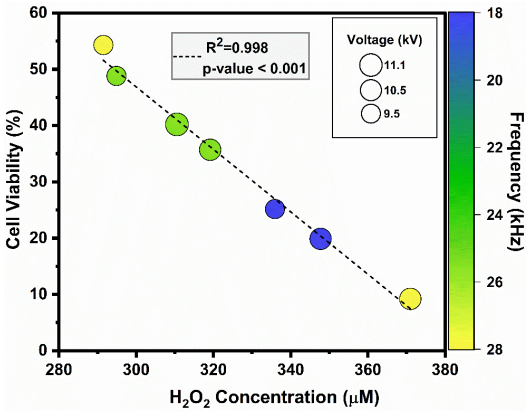

Figure 9. Linear correlation between H_2O_2 concentration and HT29 cell viability. Strong inverse relationship (R^2 = 0.995, $p < 0.001$).

of an optimal 'sweet spot' that maximizes the net production of H_2O_2 by balancing its formation and decomposition pathways. While higher voltage increases the energy for •OH formation, excessively high energy input, such as at 11.1 kV, can become counterproductive by enhancing H_2O_2 decomposition pathways (e.g., $H_2O_2 + e^- \longrightarrow 2•OH$), leading to a net decrease in its concentration (31). The optimal point at 10.5 kV and 28 kHz likely represents the ideal balance where electron energy is sufficient for robust •OH formation without excessively promoting these decomposition reactions. Frequency tuning further influenced this balance: operating at 28 kHz, beyond the system's resonant range (20–24 kHz), maximized H_2O_2 generation, likely by optimizing the pulse repetition rate to favor formation over decomposition processes (32).

The Importance of a Tunable Power Supply and Advantages over RF/DBD Sources

Plasma jets are commonly powered by a range of electrical sources, such as RF generators, pulsed DC supplies, and kHz-driven AC systems (33). The choice of a kHz AC power supply was deliberate and central to our hypothesis. The key advantage of this system over most RF power supplies (which operate at fixed frequencies) is the independent control over both voltage and frequency. Our data clearly show that the optimal H_2O_2 production is highly dependent on a specific, non-linear combination of these two parameters, a level of control fixed-parameter systems cannot offer. Furthermore, our jet configuration offers distinct advantages over DBD sources for internal tumors like CRC. DBDs are best suited for surface treatments , whereas our plasma jet produces a focused plume that can be guided down an endoscope to precisely target a tumor without invasive electrode contact (34). Critically, the therapeutic effect of our argon jet is mediated primarily by long-lived aqueous species (H_2O_2) generated at the plasma-liquid interface, which is ideal for the fluid-covered physiological environment of the colon (35). This contrasts with many DBDs whose efficacy relies more on short-lived species, intense electric fields, and UV radiation, which are less suitable when a significant liquid layer is present (35). Recent research continues to explore the unique chemical and biological effects that can be achieved with kHz-driven plasma sources, underscoring the importance of this technological approach for advancing plasma medicine (36).

The Unique Advantage and selectivity of CAP

A fundamental advantage of CAP in oncology is its inherent selectivity for cancer cells over healthy cells (37-40). This selectivity is primarily rooted in the distinct redox biology of malignant cells. Cancer cells exist in a state of chronic oxidative stress with elevated basal levels of intracellular ROS, placing them much closer to an apoptotic threshold than their healthy counterparts (41). Therefore, a precisely tuned dose of H_2O_2, like the one delivered by our system, is sufficient to overwhelm cancer cells, pushing them into apoptosis. This principle ensures that the therapeutic dose remains within the buffering capacity of adjacent healthy tissue, maintaining a high therapeutic window (42).

The strong correlation between H_2O_2 concentration and cell viability raises a critical question: why is the complexity of CAP necessary if H_2O_2 is the principal cytotoxic agent? The efficacy of CAP is not attributable to H_2O_2 alone but to the synergistic interplay of its diverse chemical and physical components (33). Although our data identify H_2O_2 as the dominant long-lived species and the most reliable predictor of cytotoxicity in this argon jet system, it does not act in isolation. It is well-established that H_2O_2 applied alone cannot fully replicate the potent and selective cytotoxicity of plasma-activated medium (PAM) (43), and even synergistic mixtures of $H_2O_2/NO_2^-/NO_3^-$ fall short of matching its biological efficacy (44, 45). Therefore, the primary advantage of our plasma jet is its ability to deliver a complex, synergistic "therapeutic cocktail," while the tunability provides precise control over the dose of its key long-lived effector, H_2O_2.

Clinical Applicability and Implications for Precision Plasma Oncology

Furthermore, the exceptionally strong linear correlation ($R^2 = 0.995$) established in this study has profound implications for the development of 'precision plasma oncology'. This relationship validates extracellular H_2O_2 concentration as a simple and reliable predictive biomarker for the biological effect of our argon plasma jet. It provides a quantitative link between tunable physical parameters (voltage, frequency) and a predictable cytotoxic outcome. This opens the door to personalized therapeutic strategies where: (1) a patient's tumor can be profiled for its specific antioxidant capacity; (2) a required cytotoxic H_2O_2 dose can be prescribed; and (3) the plasma jet can be precisely tuned to deliver that exact dose, maximizing efficacy while minimizing off-target effects. This paradigm shifts plasma therapy from a qualitative application to a quantitative, dose-controlled, and personalized treatment modality.

Limitations and Future Directions

We acknowledge several limitations in this study. Our investigation was performed on a 2D in vitro model,

which does not capture the complexity of the tumor microenvironment. While we identified H_2O_2 as the primary driver of cytotoxicity, we did not quantify other RONS or investigate detailed molecular pathways. Future work should validate these findings in more clinically relevant models, such as 3D tumor spheroids and in vivo orthotopic CRC models. Furthermore, exploring synergistic combinations of this dose-controlled plasma treatment with standard chemotherapeutics could lead to more effective, lower-dose treatment regimens (46). Finally, scaling and miniaturization of the plasma jet for endoscopic application will be crucial for its eventual clinical translation.

Conclusion

This study demonstrates that a custom-engineered, voltage- and frequency-tunable argon plasma jet is a highly effective platform for overcoming intrinsic chemoresistance in HT29 colorectal cancer cells. The central finding is the establishment of a precise, dose-controlled delivery of H_2O_2 by systematically tuning the plasma's electrical parameters. Optimal settings (10.5 kV at 28 kHz) maximized cytotoxicity, reducing cell viability to just 9.2%. This effect was underpinned by a strong linear correlation ($R^2 = 0.995$) between H_2O_2 concentration and cell death, establishing H_2O_2 as a robust predictive biomarker for therapeutic outcome. Crucially, this approach circumvents the redox resistance without sensitization, offering a blueprint for personalized, parameter-driven plasma oncology.

References

1. Roshandel G, Ghasemi-Kebria F, Malekzadeh R. Colorectal cancer: epidemiology, risk factors, and prevention. Cancers. 2024;16(8):1530.
2. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Gastroenterology Review/Przegląd Gastroenterologiczny. 2019;14(2):89-103.
3. Duineveld LA, van Asselt KM, Bemelman WA, Smits AB, Tanis PJ, van Weert HC, et al. Symptomatic and asymptomatic colon cancer recurrence: a multicenter cohort study. The Annals of Family Medicine. 2016;14(3):215-20.
4. Saoudi González N, Salvà F, Ros J, Baraibar I, Rodríguez-Castells M, García A, et al. Unravelling the complexity of colorectal cancer: Heterogeneity, clonal evolution, and clinical implications. Cancers. 2023;15(16):4020.
5. Kumar A, Gautam V, Sandhu A, Rawat K, Sharma A, Saha L. Current and emerging therapeutic approaches for colorectal cancer: A comprehensive review. World Journal of Gastrointestinal Surgery. 2023;15(4):495.
6. Keidar M, Walk R, Shashurin A, Srinivasan P, Sandler A, Dasgupta S, et al. Cold plasma selectivity and the possibility of a paradigm shift in cancer therapy. British journal of cancer. 2011;105(9):1295-301.
7. Yan D, Sherman JH, Keidar M. Cold atmospheric plasma, a novel promising anti-cancer treatment modality. Oncotarget. 2016;8(9):15977.
8. Stańczyk B, Wiśniewski M. The promising potential of cold atmospheric plasma therapies. Plasma. 2024;7(2):465-97.
9. Ratovitski EA, Cheng X, Yan D, Sherman JH, Canady J, Trink B, et al. Anti-cancer therapies of 21st century: novel approach to treat human cancers using cold atmospheric plasma. Plasma Processes and Polymers. 2014;11(12):1128-37.
10. Graves DB. Reactive species from cold atmospheric plasma: Implications for cancer therapy. Plasma Processes and Polymers. 2014;11(12):1120-7.
11. Von Woedtke T, Schmidt A, Bekeschus S, Wende K, Weltmann K-D. Plasma medicine: A field of applied redox biology. In vivo. 2019;33(4):1011-26.
12. Holanda AGA, Francelino LEC, Moura CEBd, Alves Junior C, Matera JM, Queiroz GFd. Cold Atmospheric Plasma in Oncology: A Review and Perspectives on Its Application in Veterinary Oncology. Animals. 2025;15(7):968.
13. Laroussi M, Lu X, Keidar M. Perspective: The physics, diagnostics, and applications of atmospheric pressure low temperature plasma sources used in plasma medicine. Journal of Applied Physics. 2017;122(2).
14. Weltmann KD, Polak M, Masur K, von Woedtke T, Winter J, Reuter S. Plasma processes and plasma sources in medicine. Contributions to Plasma Physics. 2012;52(7):644-54.
15. Jablonowski H, Hoffmann U, Bansemer R, Bekeschus S, Gerling T, von Woedtke T. Characterization and comparability study of a series of miniaturized neon plasma jets. Journal of Physics D: Applied Physics. 2024;57(19):195202.
16. Busco G, Fasani F, Dozias S, Ridou L, Douat C, Pouvesle J-M, et al. Changes in oxygen level upon cold plasma treatments: consequences for RONS production. IEEE Transactions on Radiation and Plasma Medical Sciences. 2017;2(2):147-52.
17. Ishaq M, Evans MD, Ostrikov KK. Atmospheric pressure gas plasma-induced colorectal cancer cell death is mediated by Nox2–ASK1 apoptosis pathways and oxidative stress is mitigated by Srx–Nrf2 anti-oxidant system. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research. 2014;1843(12):2827-37.
18. He Y, Lu F, Jiang C, Gong F, Wu Z, Ostrikov K. Cold atmospheric plasma stabilizes mismatch repair for effective, uniform treatment of diverse colorectal cancer cell types. Scientific Reports. 2024;14(1):3599.
19. Wang Y, Mang X, Li X, Cai Z, Tan F. Cold atmospheric plasma induces apoptosis in human colon and lung cancer cells through modulating mitochondrial pathway. Frontiers in Cell and Developmental Biology. 2022;10:915785.
20. Schneider C, Arndt S, Zimmermann JL, Li Y, Karrer S, Bosserhoff AK. Cold atmospheric plasma treatment inhibits growth in colorectal cancer cells. Biological chemistry. 2018;400(1):111-22.
21. Han D, Cho JH, Lee RH, Bang W, Park K, Kim MS, et al. Antitumorigenic effect of atmospheric-pressure dielectric

barrier discharge on human colorectal cancer cells via regulation of Sp1 transcription factor. Scientific reports. 2017;7(1):43081.
22. Martinet A, Miebach L, Heisterberg L, Neugebauer A, Enderle MD, Bekeschus S. Reactive Species Production and Colon Cancer Cytotoxicity of an Electrosurgical Cold Argon Plasma Device. Plasma Processes and Polymers. 2025;22(4):2400240.
23. Ginés A, Bystrup S, Ruiz de Porras V, Guardia C, Musulén E, Martínez-Cardús A, et al. PKM2 subcellular localization is involved in oxaliplatin resistance acquisition in HT29 human colorectal cancer cell lines. PloS one. 2015;10(5):e0123830.
24. Zhao Z, Zhang G, Li W. MT2A promotes oxaliplatin resistance in colorectal cancer cells. Cell Biochemistry and Biophysics. 2020;78(4):475-82.
25. Kitahara T, Haraguchi N, Takahashi H, Nishimura J, Hata T, Takemasa I, et al. Identification and characterization of CD107a as a marker of low reactive oxygen species in chemoresistant cells in colorectal cancer. Annals of surgical oncology. 2017;24:1110-9.
26. Cheraghi O, Dabirmanesh B, Ghazi F, Amanlou M, Atabakhshi-Kashi M, Fathollahi Y, et al. The effect of Nrf2 deletion on the proteomic signature in a human colorectal cancer cell line. BMC cancer. 2022;22(1):979.
27. Kim HG, Kim CW, Lee DH, Lee J-S, Oh E-T, Park HJ. Quinacrine-mediated inhibition of Nrf2 reverses hypoxia-induced 5-fluorouracil resistance in colorectal cancer. International journal of molecular sciences. 2019;20(18):4366.
28. Barrera JCA, Ondo-Mendez A, Giera M, Kostidis S. Metabolomic and lipidomic analysis of the colorectal adenocarcinoma cell line HT29 in hypoxia and reoxygenation. Metabolites. 2023;13(7):875.
29. Touil Y, Igoudjil W, Corvaisier M, Dessein A-F, Vandomme J, Monté D, et al. Colon cancer cells escape 5FU chemotherapy-induced cell death by entering stemness and quiescence associated with the c-Yes/YAP axis. Clinical cancer research. 2014;20(4):837-46.
30. Yan D, Cui H, Zhu W, Talbot A, Zhang LG, Sherman JH, et al. The strong cell-based hydrogen peroxide generation triggered by cold atmospheric plasma. Scientific reports. 2017;7(1):10831.
31. Wang H, Wandell RJ, Locke BR. The influence of carrier gas on plasma properties and hydrogen peroxide production in a nanosecond pulsed plasma discharge generated in a water-film plasma reactor. Journal of Physics D: Applied Physics. 2018;51(9):094002.
32. Harris B, Wagenaars E. The influence of pulse repetition frequency on reactive oxygen species production in pulsed He+ H2O plasmas at atmospheric pressure. Journal of Applied Physics. 2023;134(10).
33. von Woedtke T, Emmert S, Metelmann H-R, Rupf S, Weltmann K-D. Perspectives on cold atmospheric plasma (CAP) applications in medicine. Physics of Plasmas. 2020;27(7).
34. Lu X, Reuter S, Laroussi M, Liu D. Nonequilibrium atmospheric pressure plasma jets: Fundamentals, diagnostics, and medical applications: CRC Press; 2019.
35. Bekeschus S, Lin A, Fridman A, Wende K, Weltmann K-D, Miller V. A comparison of floating-electrode DBD and kINPen jet: plasma parameters to achieve similar growth reduction in colon cancer cells under standardized conditions. Plasma Chemistry and Plasma Processing. 2018;38:1-12.
36. Schweigert I, Alexandrov A, Zakrevsky D, Milakhina E, Patrakova E, Troitskaya O, et al. Mismatch of frequencies of ac voltage and streamers propagation in cold atmospheric plasma jet for typical regimes of cancer cell treatment. Journal of Physics: Conference Series. 2021;2100(1):012020.
37. Bauer G, Graves DB. Mechanisms of selective antitumor action of cold atmospheric plasma-derived reactive oxygen and nitrogen species. Plasma processes and polymers. 2016;13(12):1157-78.
38. Yan D, Horkowitz A, Wang Q, Keidar M. On the selective killing of cold atmospheric plasma cancer treatment: Status and beyond. Plasma Processes and Polymers. 2021;18(10):2100020.
39. Yan D, Talbot A, Nourmohammadi N, Sherman JH, Cheng X, Keidar M. Toward understanding the selective anticancer capacity of cold atmospheric plasma—A model based on aquaporins. Biointerphases. 2015;10(4).
40. Semmler ML, Bekeschus S, Schäfer M, Bernhardt T, Fischer T, Witzke K, et al. Molecular Mechanisms of the Efficacy of Cold Atmospheric Pressure Plasma (CAP) in Cancer Treatment. Cancers (Basel). 2020;12(2).
41. Trachootham D, Alexandre J, Huang P. Targeting cancer cells by ROS-mediated mechanisms: a radical therapeutic approach? Nature reviews Drug discovery. 2009;8(7):579-91.
42. Roshan M, Farnia P, Ghomi H, Jafari S. Apoptotic effects of cold atmospheric pressure plasma on A549 and LL/2 lung carcinoma cell lines. Scientific Reports. 2025;15(1):19567.
43. Yan D, Talbot A, Nourmohammadi N, Cheng X, Canady J, Sherman J, et al. Principles of using cold atmospheric plasma stimulated media for cancer treatment. Scientific reports. 2015;5(1):18339.
44. Girard P, Arbabian A, Fleury M, Bauville G, Puech V, Dutreix M. Synergistic effect of H2O2 and NO2 in cell death induced by cold atmospheric he plasma. Sci Rep. 2016; 6: 29098. Epub 2016/07/02. doi: 10.1038/srep29098. PubMed PMID: 27364563.
45. Kurake N, Tanaka H, Ishikawa K, Kondo T, Sekine M, Nakamura K, et al. Cell survival of glioblastoma grown in medium containing hydrogen peroxide and/or nitrite, or in plasma-activated medium. Archives of biochemistry and biophysics. 2016;605:102-8.
46. Dezhpour A, Ghafouri H, Jafari S, Nilkar M. Effects of cold atmospheric-pressure plasma in combination with doxorubicin drug against breast cancer cells in vitro and in vivo. Free Radical Biology and Medicine. 2023;209:202-10

Journal of Sciences, Islamic Republic of Iran 36(1): 61 - 67 (2025) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Fusion Reactivity of Plasma with Anisotropic Lorentzian Distribution

F. Khoshdoun[1], M. Shahmansouri[2*]

[1] *Department of Physics, Faculty of Science, Arak University, Arak, Islamic Republic of Iran*
[2] *Department of Atomic and Molecular Physics, Faculty of Physics, Alzahra University, Tehran, Islamic Republic of Iran*

Received: 9 July 2025 / Revised: 28 September 2025 / Accepted: 8 December 2025

Abstract

Anisotropic distributions and deviations from velocity equilibrium play a crucial role in plasma physics and nuclear fusion processes. The emergence of high-energy tails in non-equilibrium distributions increases the population of energetic particles, thereby enhancing the probability of quantum tunneling and, consequently, fusion reaction rates. In this work, we investigate how the velocity-space anisotropy and deviations from the equilibrium affect the optimization of the fusion yield. Specifically, we analyze non-Maxwellian distribution models, including kappa and anisotropic kappa distributions, to evaluate their impact on fusion reactivity. Our results show that anisotropic distributions outperform isotropic ones at lower temperatures, whereas isotropic distributions dominate at higher temperatures. These findings provide new insights for the design of fusion devices and contribute to improving the efficiency of fusion processes.

Keywords: Non-Maxwellian Distributions; Kappa Distribution; Fusion Reaction Rates; Velocity Anisotropy; Magnetic Confinement.

Introduction

Investigation of non-Maxwellian energy distributions in plasma physics, particularly in the context of nuclear fusion reactions, has gained increasing attention in recent years (1-10). In classical plasma systems, Maxwell-Boltzmann (MB) distributions are often inadequate for explaining the complexities of real phenomena, where particles exhibit significant deviations from equilibrium. Such deviations are observed across a wide range of environments, from space plasmas to laboratory experiments designed for controlled fusion. To better capture these effects, various non-Maxwellian models have been proposed, including kappa distributions, bi-Maxwellian distributions, and superstatistics (6). These distributions are characterized by prominent high-energy tails, which indicate an enhanced population of energetic particles and can significantly affect nuclear fusion reaction rates. For example, non-Maxwellian distributions have been associated with increased reactivity due to the greater probability of tunneling phenomena enabled by a higher number of high-energy particles (7). Additionally, velocity-space anisotropy, introduced by mechanisms such as magnetic confinement or external heating, plays a crucial role in modifying fusion yields. At plasma temperatures lower than those corresponding to the peak of the fusion cross section, anisotropic distributions generally result in higher fusion yields compared to isotropic distributions with the same thermal energy. However, as the temperature increases,

* Corresponding Author; Tel: +989339836854; Email: m.shahmansouri@alzahra.ac.ir

isotropic distributions become dominant, highlighting the complex interplay between temperature, anisotropy, and fusion efficiency (8).

The understanding of nuclear fusion has had a profound impact on the development of plasma physics. In the following, we briefly review key historical milestones in the advancement of nuclear fusion. Arthur Eddington was the first to propose that nuclear fusion could serve as the primary energy source powering stars (11). The concept of quantum tunneling was later introduced by Gamow (12), and Atkinson and Houtermans analytically investigated nuclear fusion rates in stars (13). The first artificial nuclear disintegration was achieved by Cockcroft and Walton in 1932 (14). In 1946, Arnoux was the first to patent a fusion device, which initiated experimental research on nuclear fusion the following year (15). The stellarator concept was introduced by Spitzer (16), and the first controlled fusion experiment was conducted using Scylla (17). The first generation of tokamaks, known as T-1, was constructed in the Soviet Union (18).

Over the following decades, numerous efforts were devoted to improving fusion devices and developing related physical concepts (19-21). Among the most notable achievements, JET and ITER have played leading roles: JET produced its first plasma in 1983, and conceptual design work for ITER began in 1988. In 1991, JET performed the first deuterium–tritium fusion experiment, setting a record of 16 megawatts of fusion power. More recently, JET achieved a new record with 59 megajoules of fusion energy. Very recently, the Wendelstein 7-X produced plasma with gigajoule-scale energy, while KSTAR set a record of 102 seconds of stable plasma at 10^8 ℃ in 2024 (21).

Recent research has increasingly focused on the effects of velocity-space deviations in plasmas and their influence on nuclear fusion yields. Kolmes *et al*., (13) demonstrated that velocity deviations in plasmas can significantly enhance fusion reactions, particularly at lower temperatures. They suggested that non-isotropic distributions provide superior performance at low temperatures, although this advantage diminishes as the temperature increases. Ourabah (19) examined reaction rates in systems near local equilibrium and showed that non-Maxwellian distributions can increase fusion yields, identifying specific conditions under which such distributions strongly enhance fusion activity. Furthermore, Onofrio and Sundaram (18) investigated the role of particular energy distributions and their applications in fusion devices. Their work indicated that non-Maxwellian energy distributions can improve fusion performance under certain conditions, especially in modern devices designed to exploit such distributions. Recently, Xie (26) introduced an efficient computational framework for solving plasma waves with arbitrary energy distributions, which also covers the analysis of kappa and anisotropic distributions. This method provides a useful tool for more accurate investigations of instabilities and fusion reactions in non-Maxwellian environments. Nerney (27) recently modeled anisotropic Maxwellian and Kappa distributions in Io's plasma torus, showing that distribution choice and temperature anisotropy strongly affect plasma density and temperature profiles.

The present study builds on historical advancements and recent research on non-equilibrium and anisotropic distributions-arising from external heating or magnetic confinement-to investigate how statistical particle distributions affect reaction rates in anisotropic plasma systems. The findings are expected to provide a new perspective on optimizing fusion processes and to contribute to a deeper understanding of energy transfer in complex plasma environments. This study is presented as a research article; the background material included in the introduction is intentionally concise and serves only to provide context for the new derivations and numerical analysis. In this work, we extend the standard anisotropic-kappa framework to derive new reactivity expressions by incorporating the anisotropic distribution into the fusion cross-section integral and explicitly performing the angular integration. To the best of our knowledge, this explicit angular-integral formulation for the anisotropic kappa distribution has not been reported previously. Our numerical analysis further explores how anisotropy and suprathermal tails influence reactivity across the relevant parameter ranges.

Materials and Methods

Anisotropic Kappa Distribution

The anisotropic kappa distribution is a version of the kappa distribution used in situations where the characteristics of non-equilibrium systems are distributed differently across various spatial directions. In plasma physics and astrophysics, plasma systems often exhibit anisotropic properties, meaning physical properties (e.g., temperature and density) vary across different directions (22,23).

The anisotropic kappa distribution is particularly useful for modeling such systems. The formula for the anisotropic kappa distribution in velocity coordinates $v = (v_\perp, v_\parallel)$ is as follows (24):

$$f(v) = \frac{n\Gamma(\kappa+1)}{\pi^{3/2}\theta_\perp^2\theta_\parallel\kappa^{3/2}\Gamma(\kappa-\frac{1}{2})}\left[1+\frac{1}{\kappa}\left(\frac{v_\perp^2}{\theta_\perp^2}+\frac{v_\parallel^2}{\theta_\parallel^2}\right)\right]^{-(\kappa+1)} \quad (1)$$

where n denotes the particle number density, κ is the

spectral index governing the suprathermal tail of the distribution, $\theta_\perp$ and $\theta_\parallel$ are the effective temperature parameters perpendicular and parallel to the magnetic field (or reference axis), respectively, also $v_\perp$ and, $v_\parallel$ are the corresponding velocity components. In the limit $\kappa \to \infty$, the anisotropic kappa distribution reduces to the bi-Maxwellian distribution.

In plasma physics, particularly in the analysis of particle motion in strong magnetic fields, it is common to separate the perpendicular and parallel energies with respect to the field as $E_\perp = \frac{1}{2}mv_\perp^2$ and $E_\parallel = \frac{1}{2}mv_\parallel^2$. This separation naturally allows the distribution to be expressed in terms of perpendicular and parallel energy coordinates, as follows (23):

$$f(E_\perp, E_\parallel) = \frac{n}{\pi^{3/2}\theta_\perp^2\theta_\parallel\kappa^{3/2}}\frac{\Gamma(\kappa+1)}{\Gamma(\kappa-1/2)}\left[1+\frac{\beta}{\kappa}(\beta_\perp E_\perp + \beta_\parallel E_\parallel)\right]^{-(\kappa+1)} \quad (2)$$

In Eq. (2), the parameters $\beta_\perp = 1/\beta\theta_\perp$ and $\beta_\parallel = 1/\beta\theta_\parallel$ represent the inverse effective temperatures in the perpendicular and parallel directions, respectively; and β as a normalization factor is given by $0.1(keV)^{-1}$ refers to a typical value of deuterium-tritium reaction (28).

To obtain the distribution function as a function of the total kinetic energy, one must integrate the velocity-space distribution over all angles in spherical coordinates. For an anisotropic, axis- preferred distribution (e.g., aligned with the magnetic field vector), it is convenient to adopt spherical coordinates, with $v_\parallel = vcos\theta$, $v_\perp = vsin\theta$ and $E = \frac{1}{2}mv^2$. The number of particles with energy between E and $E + dE$ is then given by the following integral expression (23):

$$f(E) = 4\pi \int f(v)\delta\left(\frac{1}{2}mv^2 - E\right)v^2 dv \quad (3)$$

In Eq. (3), $f(E)$ denotes the single-particle energy distribution obtained by integrating the velocity-space formulation. The angle θ is the polar angle measured relative to the magnetic-field (parallel) direction. This transformation ensures that the anisotropic kappa distribution can be expressed consistently in terms of the total particle energy. For normalization, the integration is restricted to the region where $E = \frac{1}{2}mv^2$. In spherical velocity coordinates (relative to the parallel axis), the volume element is $d^3v = v^2 sin\theta dv d\varphi d\theta$. Assuming azimuthal symmetry about φ (rotation around the parallel axis), the integral over φ contributes only a factor of 2π, consistent with the azimuth-integrated forms.

The result (without delving into the full details of integration) is as follows (23):

$$f(E) = \frac{\Gamma(\kappa+1)}{\kappa^{3/2}\Gamma(\kappa-\frac{1}{2})}\underbrace{\frac{n}{\pi^{3/2}\theta_\perp^2\theta_\parallel}}\left(\frac{E}{m}\right)^{\frac{3}{2}}\int_0^\pi\left[1+\frac{\beta E}{\kappa}(\beta_\perp sin^2\theta + \beta_\parallel cos^2\theta)\right]^{-(\kappa+1)} sin\theta d\theta \quad (4)$$

In Eq. (4), $f(E)$ denotes the azimuthally integrated anisotropic kappa distribution expressed as a function of the total particle energy. This compact form, obtained by carrying out the angular integrations, provides a convenient representation of the energy distribution for subsequent evaluation of fusion reactivity. The resulting $f(E)$ is fully consistent with the azimuth-independent anisotropic-kappa formulation.

In the following, we derive an explicit expression for the fusion reactivity in anisotropic, non-thermal plasmas. The tunneling cross section, formulated in terms of the anisotropic kappa velocity distribution $f(v)$, takes the form (25):

$$\langle\sigma v\rangle = \int_0^\infty\int_0^\pi\left[1+\frac{\beta E}{\kappa}(\beta_\perp sin^2\theta + \beta_\parallel cos^2\theta)\right]^{-(\kappa+1)} T(E)E^{-1/2}\, d\theta dE \quad (5)$$

In Eqs. (5)-(8), $\langle\sigma v\rangle$ denotes the fusion reactivity, defined as the velocity-averaged product of the fusion cross section $\sigma(E)$ and the relative velocity v of the reacting particles. Here, $\sigma(E)$ is the energy-dependent nuclear fusion cross section, while v represents the relative speed between the interacting nuclei. This quantity serves as the fundamental link between microscopic nuclear properties and macroscopic reaction rates in plasmas.

The angular integration over $d\theta$ explicitly incorporates anisotropy arising from temperature differences in the perpendicular and parallel directions. The transmission coefficient $T(E)$ is typically modeled with either exponential or power-law dependence:

$$T(E) = \begin{cases}\left(\frac{E}{E_0}\right)^\alpha & for\ E < E_0, \\ 1 & for\ E > E_0.\end{cases} \quad (6)$$

where E_0 refers to threshold energy that serves as a modeling parameter that defines the boundary between the gradual growth region and saturation region of the tunneling coefficient; α represents role of a convexity parameter varies between 0 and ∞, as for $E < E_0$, we have $T(E)$ concave (convex) if α < 1 (if α > 1). This formulation incorporates the effect of anisotropic temperatures ($\beta_\perp$ and $\beta_\parallel$) on reactivity calculations.

Reactivity ($\delta R_{a\kappa}$) of fusion

To quantify the enhancement of reactivity in the anisotropic kappa distribution relative to the Maxwell–Boltzmann case, we evaluate and compare the reactivity

within two distinct frameworks, as follows:

$$\delta R_{a\kappa} = \langle \sigma v \rangle_{a\kappa} - \langle \sigma v \rangle_{MB} \quad (7)$$

Substituting the anisotropic kappa distribution from Eq. (4) into δR_κ we obtain:

$$\delta R_{a\kappa} = \int_0^\infty \int_0^\pi \left(\left[1 + \frac{2\beta E}{\kappa} (sin^2\phi\, \beta_\perp + cos^2\phi\, \beta_\parallel) \right]^{-(\kappa+1)} - p_{MB}(E) \right) T(E) E^{-1/2}\, d\phi dE \quad (8)$$

where:

$$p_{MB}(E) = \sqrt{\frac{\beta}{4\pi}} E^{-1/2} \exp(-\beta E) \quad (9)$$

The angular part of integration in Eq. (8) is evaluated using the Gauss–Kronrod algorithm to compute single-variable integrals (8).

Results and Discussion

To investigate the influence of anisotropy and deviations from equilibrium, we numerically analyze the dependence of fusion reactivity on the anisotropic parameters ($\beta_\perp, \beta_\parallel$) and the spectral index κ. Figure 1, illustrates pronounced differences between the anisotropic and standard kappa distributions. At high energies (above $\beta E = 10^1$), the particle density in the anisotropic case is significantly reduced; for instance, at $\beta E = 10^2$, the value of $f(E)$ is nearly an order of magnitude lower than in the standard kappa distribution. This suppression of suprathermal tails enhances plasma stability and reduces energy losses. Here, plasma stability refers specifically to kinetic regimes where suprathermal particles do not strongly drive wave–particle instabilities and transport-related energy losses remain minimized. More precisely, stability is defined as the ability of the anisotropic kappa distribution to remain self-consistent without exciting mirror, whistler, or firehose modes, while simultaneously reducing collisional and turbulent losses. Our numerical results confirm this definition; as the parameter ranges that enhance reactivity (low κ and moderate $T_\perp/T_\parallel$) coincide with conditions where suprathermal particles are less destabilizing.

At low energies (below $\beta E = 10^{-1}$), the anisotropic distribution exhibits higher particle densities. For example, at $\beta E = 10^{-2}$, $f(E)$ is nearly an order of magnitude greater than in the standard kappa distribution. This enhancement is particularly favorable for fusion reactions at low temperatures. In the intermediate regime ($10^{-1} < \beta E < 10^1$), the anisotropic kappa distribution is more compressed and has a shallower slope compared to the standard case, highlighting anisotropy as a potential mechanism for controlling energy distribution in plasmas.

The parameters $\beta_\parallel$ and $\beta_\perp$ independently regulate transverse and longitudinal energies. Larger values $\beta_\parallel =$

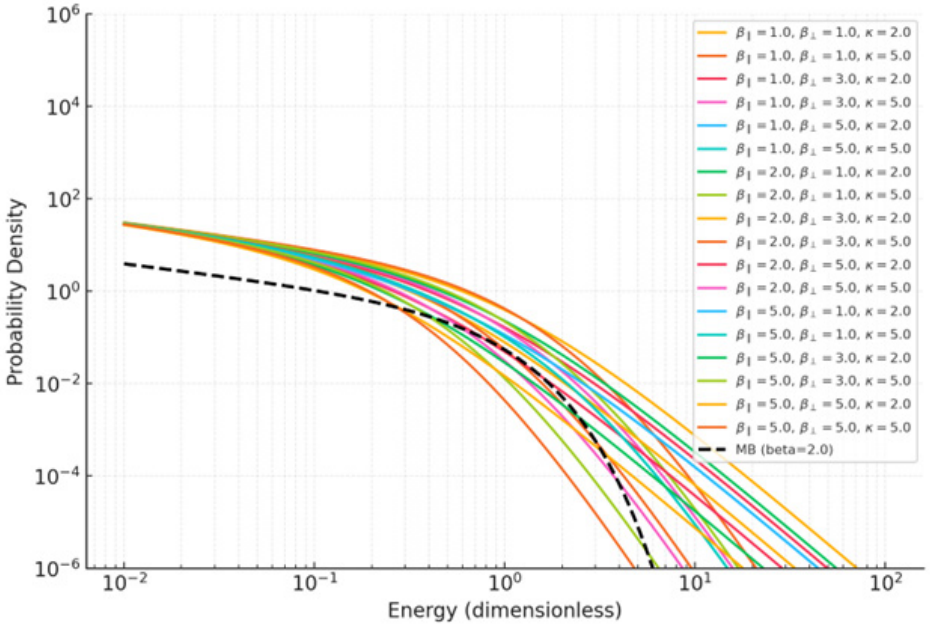

Figure 1. Anisotropic Kappa distributions for the various values of β and κ, as a function of energy. Colored lines indicate different values of $\beta_\parallel$ and $\beta_\perp$. The dashed black line represents the Maxwell-Boltzmann distribution for $\beta = 2.0$ as a reference.

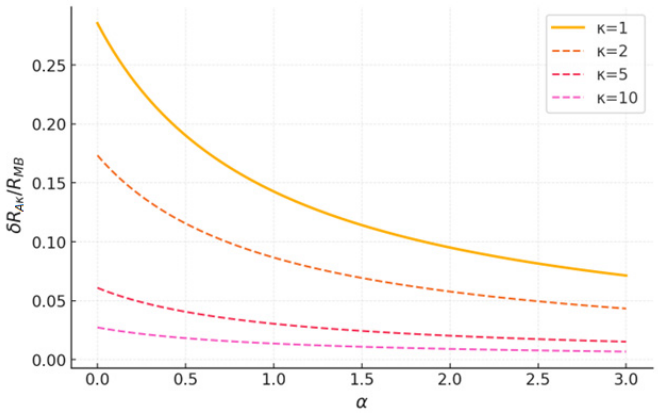

Figure 2. Variation of $\delta R_{A\kappa}/R_{MB}$ as a function of α for different values of spectral parameter κ.

5.0 and $\beta_\perp = 5.0$ shift the anisotropic kappa distribution closer to the Maxwell–Boltzmann (MB) form, although the extended high-energy tails characteristic of kappa distributions remain. The dashed black curve in Figure 1 represents the MB distribution for $\beta = 2.0$. At high energies ($\beta E > 10^1$), this curve lies well below the kappa distributions and lacks suprathermal tails, but at intermediate and low energies, the anisotropic kappa distribution converges toward *MB*. In summary, the anisotropic kappa distribution reduces high-energy tails while increasing particle densities at low energies, thereby providing a more favorable balance than the standard kappa distribution. These features enhance plasma stability, lower energy losses, and optimize conditions for fusion reactivity, making the anisotropic kappa framework particularly attractive for next-generation plasma confinement systems.

Figure 2 examines the variation of $\delta R_{a\kappa}/R_{MB}$ with α. It can be seen that as α increases, the reactivity decreases. For $\kappa = 1$, the ratio drops from approximately 0.25 at $\alpha = 0$ to 0.05 at $\alpha = 3$. For $\kappa = 2$, the decline is from about 0.15 to 0.03, while for $\kappa = 5$ and $\kappa = 10$ the initial values (≈ 0.08 and ≈ 0.05, respectively) decrease to 0.015 and 0.01 at $\alpha = 3$. These results show that small κ values ($\mathrm{e.g.}, \kappa = 2$) yield higher reactivity at low energies and small α, but fall off more steeply with increasing α. In contrast, larger κ values ($\mathrm{e.g.}, \kappa = 10$) exhibit more stable behavior and perform better at higher temperatures. Increasing $\beta_\perp$ generally enhances overall reactivity, particularly at small κ. This analysis underscores the importance of optimizing κ, α, and $\beta_\perp$ to maximize reactivity under diverse plasma conditions. To compute the reactivities the experimental cross sections parametrized in Ref. (28) has been used by integrating the reactivity over the energy validated intervals.

Figure 3 further illustrates the dependence of relative reactivity ($\delta R_{a\kappa}/R_{MB}$) on α for different $\beta_\parallel$ values. With $\beta_\parallel = 5.0$, the reactivity at $\alpha = 0$ is about 2.5 for $\kappa = 2$ and 1.5 for $\kappa = 10$. As α increases, the values decrease to ≈ 0.5 and ≈ 0.8, respectively, at $\alpha = 2.5$. Thus, small κ yields higher reactivity at low energies but declines more steeply. For $\beta_\parallel = 10$, the reactivity rises: at $\alpha = 0$, the values are ≈ 3.5 for $\kappa = 2$ and ≈ 1.8 for $\kappa = 10$. By $\alpha = 2.5$, they reduce to ≈ 1.0 and ≈ 0.9, respectively. This trend confirms that smaller κ values perform better at low energies but are more sensitive to increasing α. For $\beta_\parallel = 20$, the reactivity at $\alpha = 0$ reaches ≈ 5.5 for $\kappa = 2$ and ≈ 2.2 for $\kappa = 10$. As α increases, the values decrease to ≈ 1.5 and ≈ 1.2, respectively. Again, $\kappa = 2$ delivers stronger performance at low energies but declines more sharply than $\kappa = 10$.

Conclusion

This study highlights the critical role of non-Maxwellian energy distributions and anisotropy in plasma physics, with particular emphasis on kappa distributions and their comparison to the Maxwell–Boltzmann (MB) case. The main conclusions are as follows:

In anisotropic plasmas, temperature differences between parallel ($\beta_\parallel$) and perpendicular ($\beta_\perp$) directions strongly shape the particle energy distribution.

At low α, the contribution of suprathermal particles to fusion is negligible, and the fusion rate ($R_{A\kappa}$) is nearly identical to the MB result (R_{MB}).

As α increases, the extended high-energy tails of kappa distributions substantially enhance tunneling probabilities and fusion rates, particularly for small κ.

When $\beta_\parallel > \beta_\perp$, anisotropy further amplifies the role of high-energy particles in the perpendicular direction,

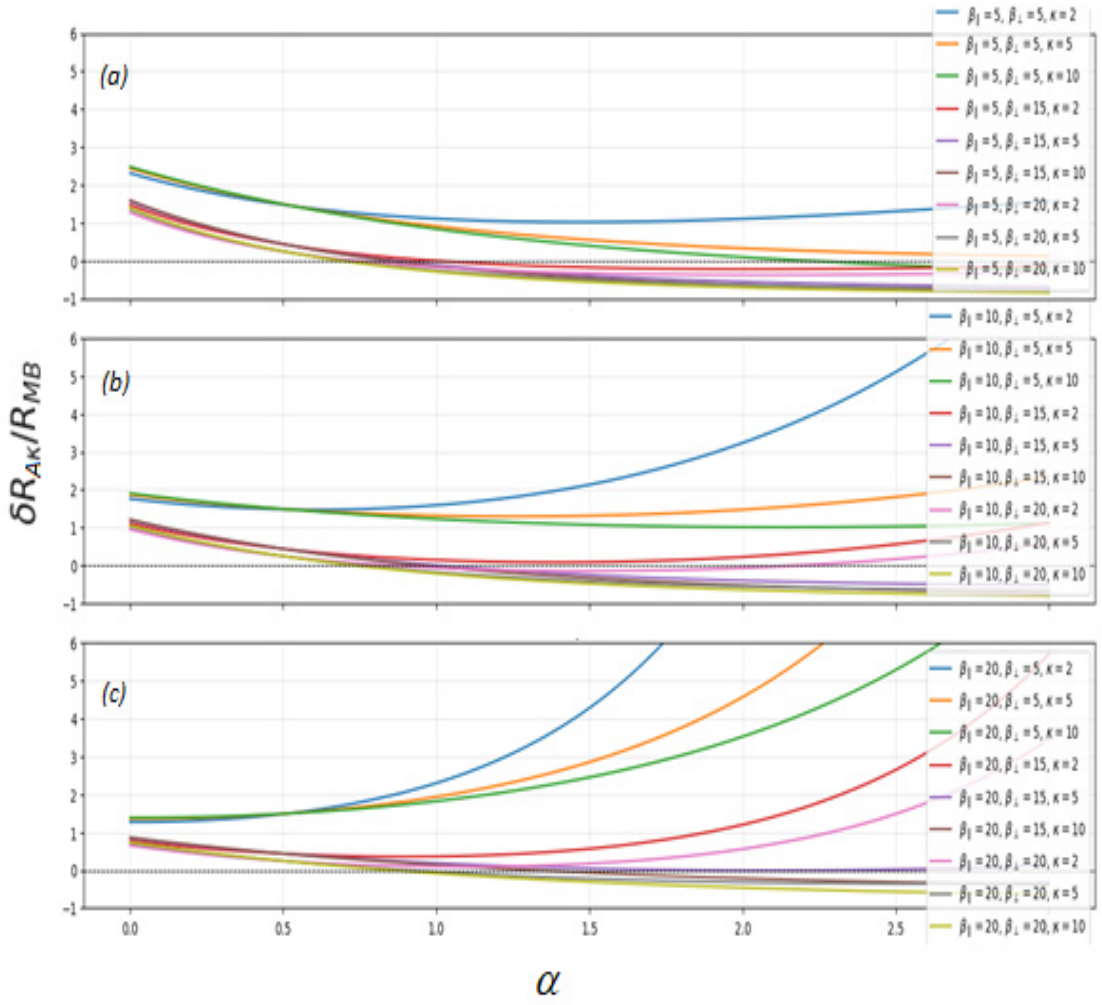

Figure 3. Comparison of the variation of $\delta R_{\kappa}/R_{MB}$ for different values of κ and $\beta_{\perp}$ across the three values of (a) $\beta_{\parallel}=5$, (b) $\beta_{\parallel}=10$ and (c) $\beta_{\parallel}=20$.

yielding higher fusion rates compared to MB.

The temperature ratio determines the dominant contribution to the overall fusion rate; increasing $T_{\perp}$ under suitable conditions significantly boosts reactivity.

Higher α values magnify sensitivity to high-energy particles, underscoring the advantage of kappa distributions for improving fusion efficiency.

Overall, kappa distributions-together with controlled anisotropy-offer far greater flexibility than MB in modeling the high-temperature plasmas. They not only enhance fusion rates but also enable more effective use of energetic particles under tailored confinement conditions. In magnetic confinement systems, external heating and temperature anisotropy can be exploited to maximize yields, while in inertial confinement, engineered plasma environments benefit from non-Maxwellian statistics.

By fine-tuning key parameters (κ , α , β), fusion efficiency can be significantly improved. These findings point toward practical pathways for optimizing confinement strategies and advancing the development of sustainable and clean nuclear fusion energy.

References

1. Luo L, Ma M, Yang Y, Zhao H. Case Report: Genetic 1. Borhanian J, Shahmansouri M. Spherical electron acoustic solitary waves in plasma with suprathermal

electrons. Astrophys Space Sci. 2012;342:401.
2. Shahmansouri M, Borhanian J. Spherical Kadomtsev–Petviashvili solitons in a suprathermal complex plasma. Commun Theor Phys. 2013;60:227.
3. Shahmansouri M, Tribeche M. Large amplitude dust ion acoustic solitons and double layers in dusty plasmas with ion streaming and high-energy tail electron distribution. Commun Theor Phys. 2014;61:377.
4. Ainejad H, Mahdavi M, Shahmansouri M. Effects of superthermal electrons and negatively (positively) charged dust grains on dust-ion acoustic wave modulation. Eur Phys J Plus. 2014;129(5):99.
5. Shahmansouri M, Tribeche M. Dust acoustic shock waves in suprathermal dusty plasma in the presence of ion streaming with dust charge fluctuations. Astrophys. Space Sci. 2013;343:251.
6. Bethe HA. Energy production in stars. Phys Rev. 1939;55(5):434–56. doi:10.1103/PhysRev.55.434.
7. Bethe HA, Critchfield CL. The formation of deuterons by proton combination. Phys Rev. 1938;54(4):248–54. doi:10.1103/physrev.54.248.
8. Calió F, Gautschi W, Marchetti E. On computing Gauss-Kronrod quadrature formulae. Math Comput. 1986;47(176):639–50.
9. Crilly AJ, Appelbe BD, Mannion OM, Taitano W, Chittenden JP. Constraints on ion velocity distributions from fusion product spectroscopy. Nucl Fusion. 2022;62(12):126015.
10. Daniel HP, Moya PS, Zenteno-Quinteros B, Colmenares GA. Magnetic spectra comparison for kappa-distributed whistler electron fluctuations. Astrophys J. 2024;970(1):132.
11. Eddington AS. The internal constitution of the stars. Science. 1920;52(1341):233–40.
12. Hendry J, Lawson J. Fusion research in the UK 1945–1960. AEA Technology; 1993.
13. Kolmes EJ, Mlodik ME, Fisch NJ. Fusion yield of plasma with velocity-space anisotropy at constant energy. Princeton University, Department of Astrophysical Sciences; 2021.
14. Li K, Liu ZY, Yao YL, Zhao ZH, Dong C, Li D, et al. Modification of the fusion energy gain factor in magnetic confinement fusion due to plasma temperature anisotropy. Nucl Fusion. 2022;62(12):126015.
15. Miyamoto K. Plasma physics for controlled fusion. Berlin Heidelberg: Springer-Verlag; 2016.
16. Yang S, et al. Tailoring tokamak error fields to control plasma instabilities and transport. Nat Commun. 2024;15:1275. doi:10.1038/s41467-024-45454-1.
17. Nimtz G, Clegg B. Compendium of quantum physics. Berlin: Springer; 2009. p. 799–802. ISBN 978-3-540-70622-9.
18. Onofrio R, Sundaram B. Relationship between nonlinearities and thermalization in classical open systems: The role of the interaction range. Phys Rev E. 2024;101(1):123–34.
19. Ourabah K. Reaction rates in quasiequilibrium states. Phys Rev E. 2024;111:034115.
20. Rebut P-H. The JET preliminary tritium experiment. Plasma Phys Control Fusion. 1992;34(13):1749–58. doi:10.1088/0741-3335/34/13/002.
21. Smirnov VP. Tokamak foundation in USSR/Russia 1950–1990. Nucl Fusion. 2009;50(1):014003. doi:10.1088/0029-5515/50/1/014003.
22. Squarer BI, Presilla C, Onofrio R. Enhancement of fusion reactivities using non-Maxwellian energy distributions. Phys Rev E. 2024;101(1):123–34.
23. Livadiotis G, Nicolaou G, Allegrini F. Anisotropic kappa distributions. I. Formulation based on particle correlations. Astrophys J. 2021;253:16.
24. Shahmansouri M, Lee MJ, Jung YD. Anisotropic temperature effects on Landau damping in Kappa-Maxwellian astrophysical plasmas. Astropart Phys. 2020;130:102449. doi:10.1016/j.astropartphys.2020.102449.
25. Atkinson RdE, Houtermans FG. Zur Frage der Aufbaumöglichkeit der Elemente in Sternen. Z Phys. 1929;54(9–10):656–65. doi:10.1007/BF01341595.
26. Xie H. Efficient framework for solving plasma waves with arbitrary distributions. Phys. Plasmas 2025;32:060702.
27. Nerney EG. Diffusive equilibrium: Modeling anisotropic Maxwellian and Kappa field line distributions in Io's plasma torus using multi-fluid and kinetic approaches. J Geophys Res Space Phys. 2025. doi:10.22541/essoar.175009109.95959011/v1
28. Bosch HS, Hale, GM. Improved formulas for fusion cross-sections and thermal reactivities. Nucl. Fusion 1992;32:611.

PERSIAN TRANSLATION OF ABSTRACTS

چکیده‌های فارسی

Genetic Analysis of Y-STRs in Two Iranian Sub-Populations

M. Ghaderi-Zefrehei[1], F. Rafeie[2], A. Alipour Tabrizi[3], Z. Baratieh[4], S. K. Fotouhi[4], J. Smith[5], M. Muhaghegh Dolatabady[1*]

[1] Department of Animal Science, Faculty of Agricultural Sciences, Yasouj University, Yasouj, Islamic Republic of Iran
[2] Department of Agricultural Biotechnology, Faculty of Agricultural Sciences, University of Guilan, Rasht, Islamic Republic of Iran
[3] Molecular Genetics Department, Mashhad Forensic Medical Center, Mashhad, Islamic Republic of Iran
[4] Forensic Medical Research Center, National Forensic Medical Organization, Tehran, Islamic Republic of Iran
[5] The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK
* Email: mmuhaghegh@yu.ac.ir

تجزیه و تحلیل ژنتیکی Y-STRها در دو زیر جمعیت ایرانی

مصطفی قادری-زفرقهی[۱]، فرجاد رفیعی[۲]، آرش علیپور تبریزی[۳]، زهره براتیه[٤]، سید کمال فتوحی[٤]، ژاکلین اسمیت[٥]، مصطفی محقق دولت آبادی[۱*]

[۱] گروه علوم دامی، دانشکده علوم کشاورزی، دانشگاه یاسوج، یاسوج، جمهوری اسلامی ایران
[۲] گروه بیوتکنولوژی کشاورزی، دانشکده علوم کشاورزی، دانشگاه گیلان، رشت، جمهوری اسلامی ایران
[۳] گروه ژنتیک مولکولی، مرکز پزشکی قانونی مشهد، مشهد، جمهوری اسلامی ایران
[٤] مرکز تحقیقات پزشکی قانونی، سازمان پزشکی قانونی کشور، تهران، جمهوری اسلامی ایران
[٥] مؤسسه روزلین و دانشکده سلطنتی دامپزشکی ، دانشگاه ادینبرو، ایستر بوش، میدلوثیان، EH25 9RG، انگلستان

چکیده

این مطالعه، تجزیه و تحلیل ژنتیکی جامعی از ۱۷ جایگاه تکرار پشت سر هم کوتاه کروموزوم Y (Y-STR) در دو زیرجمعیت ایرانی از استان‌های فارس (n=109) و اصفهان (n=180) ارائه می‌دهد. با توجه به ماهیت اختصاصی وجود در جنس نر، فقدان نوترکیبی و وراثت تک والدی کروموزوم Y ، Y-STRها به عنوان نشانگرهای مولکولی ارزشمندی در تحقیقات پزشکی قانونی، تبارشناسی و ژنتیک جمعیت عمل می‌کنند. در این مطالعه، جایگاه‌های مورد بررسی شامل DYS19 ، DYS385a/b، DYS389I/II، DYS390، DYS391، DYS392، DYS393، DYS437، DYS438، DYS439، DYS448، DYS456، DYS458، DYS63و Y-GATA-H4 بودند. نتایج نشان داد که جایگاه DYS385a/b ، با ۱۱ آلل مجزا شناسایی شده و میانگین تعداد آللها به ترتیب ۶.۲۹ و ۵.۸۸ در گروه‌های فارس و اصفهان بیشترین تنوع آللی را در هر دو جمعیت نشان داد. در مقابل، گروه فارس کمترین تنوع آللی (۳ آلل) را در جایگاه‌های DYS439 و Y-GATA-H4 نشان داد، در حالی که جمعیت اصفهان کمترین تنوع (۴ آلل) را در DYS19 و DYS439 نشان داد. تجزیه و تحلیل هاپلوتایپ نشان داد که نرخ اشتراک درون جمعیتی در فارس ۲/۷۵ درصد و در اصفهان ۱۰/۰ درصد است، و در مجموع ۸/۳ درصد همپوشانی هاپلوتایپ در مجموعه داده‌های ترکیبی ۲۸۹ فرد مشاهده شد. هر دو جمعیت مقادیر تنوع هاپلوتایپ بالایی نزدیک به ۰/۹۹را نشان دادند که نشان‌دهنده تنوع ژنتیکی قابل توجه است. ظرفیت تمایز هاپلوتیپ در بین جمعیت‌ها متفاوت بود و برای فارس ۰/۹۷۲۵، برای اصفهان ۰/۸۵۱۹ و برای کل مجموعه نمونه ۰/۹۱۷۰ ثبت شد. تمایز جمعیت‌ها با استفاده از معیارهای FST و RST جفتی ارزیابی شد که واگرایی ژنتیکی قابل توجهی را بین گروه‌های فارس و اصفهان تأیید کرد (FST = 0.00743، p < 0.001 ؛ RST = 0.0106، p < 0.01). این یافته‌ها تمایز ژنتیکی دو زیرجمعیت را برجسته می‌کند. این مطالعه بر لزوم تحقیقات بیشتر با در نظر گرفتن چندشکلی‌های تک نوکلئوتیدی کروموزوم Y (Y-SNPs)، اندازه نمونه‌های بزرگتر و اطلاعات اجدادی اضافی برای افزایش درک ساختار ژنتیکی و تاریخچه جمعیتی در جمعیت‌های ایرانی تأکید می‌کند.

واژه های کلیدی: هاپلوتیپ؛ تنوع ژنتیکی؛ نشانگر Y-STR

Antiretroviral Therapy Among HIV-Infected Pregnant Women on Their Offspring

M. Montazeri[1], M.A. Davarpanah[2], M. Montazeri[3], L. Davoodi[4*]

[1] Razi Clinical Researches Development, Mazandaran University of Medical Sciences, Sari, Islamic Republic of Iran
[2] Department of Infectious and Tropical Diseases, Faculty of Medicine, Shiraz University of Medical Sciences, Shiraz, Islamic Republic of Iran
[3] Sarem Fertility and Infertility Research Center (SAFIR), Sarem Women's Hospital, Tehran, Islamic Republic of Iran
[4] Department of Infectious Diseases, Antimicrobial Resistance Research Center, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Islamic Republic of Iran
* Email: lotfdavoodi@yahoo.com

تأثیر درمان ضد رتروویروسی در زنان باردار مبتلا به HIV بر روی فرزندانشان

مریم منتظری[۱]، محمد علی داورپناه[۲]، مریم منتظری[۳]، لطف اله داوودی[۴*]

[۱] توسعه تحقیقات بالینی رازی، دانشگاه علوم پزشکی مازندران، ساری، جمهوری اسلامی ایران
[۲] گروه بیماری‌های عفونی و گرمسیری، دانشکده پزشکی، دانشگاه علوم پزشکی شیراز، شیراز، جمهوری اسلامی ایران
[۳] مرکز تحقیقات باروری و ناباروری صارم، بیمارستان صارم، تهران، جمهوری اسلامی ایران
[۴] گروه بیماری‌های عفونی، مرکز تحقیقات مقاومت میکروبی، دانشکده پزشکی، دانشگاه علوم پزشکی مازندران، ساری، جمهوری اسلامی ایران

چکیده

خطر انتقال عفونت ویروس نقص ایمنی انسانی (HIV) از مادر به فرزند تقریباً ۳۰٪ است. با این حال، داروهای ضد رتروویروسی می‌توانند انتقال از مادر به کودک را به کمتر از ۲٪ کاهش دهند. این مطالعه به منظور تعیین تأثیر درمان ضد رتروویروسی در زنان مبتلا به HIV و پیامدهای مادری و نوزادی آن در ایران طراحی شده است. این مطالعه گذشته‌نگر با بررسی داده‌های مادر و نوزاد در شهر شیراز بین سال‌های ۲۰۰۶ تا ۲۰۱۲ طراحی شده است. زنان باردار مبتلا به HIV به دو گروه مداخله (دریافت کننده شیمی‌درمانی پروفیلاکسی) و کنترل (بدون دریافت هیچ درمانی) تقسیم شدند. اطلاعات مادر و نوزاد استخراج و ثبت شد. داده‌ها وارد نرم‌افزار SPSS شده و مورد تجزیه و تحلیل قرار گرفتند. انتقال مادر به کودک در گروه مداخله ۲.۹٪ در مقایسه با ۱۵.۸٪ در گروه کنترل بود (OR=0.01, 95% CI: 0.002-0.125, p<0.0001) میزان ابتلا به HIV در نوزادان پسر (OR=2.76, CI 95% 2.213-3.327)، زایمان طبیعی (OR=3.78, CI 95% 3.140-4.409) و شیردهی (OR=26, CI 95% 7.87-85.90) به طور قابل توجهی بالاتر بود. مداخله درمانی به طور قابل توجهی انتقال HIV از مادران آلوده به نوزادان را کاهش می‌دهد. با این حال، میزان انتقال عمودی در ایران علیرغم مداخلات درمانی هنوز بالاتر از میزان گزارش شده در کشورهای توسعه یافته است و اقدامات پیشگیرانه بیشتری ضروری به نظر می‌رسد.

واژه‌های کلیدی: شیمی درمانی؛ درمان؛ بارداری؛ HIV؛ انتقال مادر به کودک

A New Approach for Normalizing Continuous Data, Applicable in Parametric and Nonparametric Continuous Studies

M.M Saber[1], M. Taghipour[2*], M. Salehi[2], H.M Yousof[3]

[1] Department of Statistics, Higher Education Center of Eghlid, Eghlid, Islamic Republic of Iran.
[2] Department of Statistics, Faculty of Sciences, University of Qom, Qom, Islamic Republic of Iran.
[3] Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Benha University, Egypt.
** Email m.taghipour@qom.ac.ir*

یک رویکرد جدید برای نرمال سازی داده‌های پیوسته، قابل اجرا در مطالعات پیوسته پارامتریک و ناپارامتریک

محمدمهدی‌صابر[۱] مهردادتقی‌پور[۲*]، محسن‌صالحی[۳]، هیثم یوسف[٤]

[۱] گروه آمار، مرکزآموزش عالی اقلید، اقلید، جمهوری اسلامی ایران
[۲] گروه آماردانشکده‌علوم‌پایه، دانشگاه‌قم، قم، جمهوری اسلامی ایران
[۳] گروه آمار، ریاضی و بیمه، دانشکده‌علوم‌پایه، دانشگاه بنها، مصر

چکیده

نرمال بودن داده ها یک فرض اساسی در بسیاری از استنباط‌های آماری است و نبودن این فرض می‌تواند بر استنباط و نتایج بدست آمده تأثیر بگذارد. در این مقاله، یک روش جدید برای نرمال‌سازی داده‌ها ارائه می‌دهیم. این روش بر اساس یک نسخه جدیدی از تابع توزیع تجربی ایجاد شده است. روش فوق مشکلات و معایبی که در نرمال‌سازی داده‌ها بوجود می‌آید را ندارد. علاوه بر این، روش ما برگشت‌پذیر است، و اجازه می‌دهد تا داده‌های نرمال شده به راحتی به شکل اصلی تبدیل شوند. عملکرد روش پیشنهادی از طریق چند نمونه از جمله چندین مجموعه داده Covid-19 بررسی شده است.

واژه‌های کلیدی: فرض نرمال؛ ، تبدیل باکس-کاکس؛ تبدیل یئو جانسون؛ تابع توزیع تجربی

Hybrid Prediction Models for Suicide Mortality Levels in Iranian Provinces: Spatial Econometrics vs. Random Forests

M. A. Farzammehr1*, Geoffrey J. McLachlan[2]

[1]*Judiciary Research Institute, Tehran, Islamic Republic of Iran.*
[2]*Department of Mathematics, University of Queensland, Brisbane, Australia.*
* *Email: m.farzammehr@jri.ac.ir*

مدل‌های ترکیبی پیش‌بینی سطوح مرگ‌ومیر ناشی از خودکشی در استان‌های ایران: اقتصادسنجی فضایی در برابر جنگل تصادفی

محدثه‌السادات فرزام‌مهر[۱]*، جف مک لاکلن[۲]

[۱] پژوهشگاه قوه قضاییه، تهران، جمهوری اسلامی ایران
[۲] گروه ریاضی، دانشگاه کوئینزلند، بریزبن، استرالیا

چکیده

با گسترش استفاده از روش‌های پیشرفته یادگیری ماشین مانند جنگل تصادفی، درک اهمیت عوامل فضایی در این مدل‌ها بیش از پیش ضرورت یافته است. این پژوهش رویکردی نوآورانه ارائه می‌دهد که در آن مدل‌های طبقه‌بندی جنگل تصادفی با افزودن متغیرهای دارای وقفه فضایی، مطابق با ساختار مدل‌های اقتصادسنجی پانلی فضایی، توسعه یافته‌اند. عملکرد نسبی این مدل‌ها در مقایسه با روش‌های کلاسیک رگرسیون فضایی و غیرفضایی برای پیش‌بینی نرخ مرگ‌ومیر ناشی از خودکشی در ۳۱ استان ایران و با استفاده از داده‌های سال‌های ۲۰۱۱ تا ۲۰۲۱ ارزیابی شده است. نتایج نشان می‌دهد که مدل جنگل تصادفی فضایی که شامل پارامترهای وقفه فضایی است، دقت چشمگیر ۸۹.۱۹ درصدی در پیش‌بینی سطوح مرگ‌ومیر ناشی از خودکشی دارد؛ در حالی که مدل‌های کلاسیک اقتصادسنجی فضایی تنها ۴۶.۵۱ درصد و مدل‌های جنگل تصادفی غیرفضایی ۲۷.۰۳ درصد دقت دارند. ضمن تأکید بر کارآمدی مدل جنگل تصادفی فضایی با پارامترهای وقفه فضایی، این مطالعه همچنان به اهمیت مدل‌های کلاسیک اقتصادسنجی فضایی در پیش‌بینی نرخ مرگ‌ومیر ناشی از خودکشی اشاره می‌کند. یافته‌های این پژوهش بینش‌های ارزشمندی درباره تعامل میان ملاحظات فضایی و مدل‌سازی پیش‌بینی‌کننده ارائه می‌دهد و راهنمایی‌های مهمی برای پژوهشگران در انتخاب مدل‌های مناسب تحلیل داده‌های فضایی فراهم می‌سازد.

واژه‌های کلیدی: مدل‌های اقتصادسنجی فضایی؛ روش‌های یادگیری ماشین؛ جنگل تصادفی؛ ایران

Precision Tuning of a kHz-Driven Argon Plasma Jet Enables Dose-Controlled H_2O_2 Delivery to Overcome Chemoresistance in Colorectal Cancer

A. Eftekharinasab*, H. Mehdian, A. Hasanbeigi

Department of Physics and Institute for Plasma Research, Kharazmi University, Tehran, Islamic republic of Iran

**Email: eftekharinasab@gmail.com*

تنظیم دقیق جت پلاسمای آرگون با فرکانس کیلوهرتز: راهکاری برای رساندن H_2O_2 با دوز کنترل‌شده و غلبه بر مقاومت شیمی‌درمانی در سرطان کولورکتال

افشین افتخارنسب*، علی حسن بیگی، حسن مهدیان

گروه فیزیک و پژوهشکده پلاسما، دانشگاه خوارزمی، تهران، جمهوری اسلامی ایران

چکیده

سرطان کولورکتال یکی از چالش‌های عمده درمانی به‌شمار می‌آید که بخش قابل‌توجهی از آن ناشی از مکانیسم‌های مقاومتی قوی در برابر شیمی‌درمانی، از جمله فعال‌سازی مسیرهای آنتی‌اکسیدانی است. اگرچه پلاسما سرد اتمسفری به‌عنوان یک روش نوین ضدسرطان مطرح شده، اما اثربخشی آن ممکن است تحت تأثیر این دفاع‌های سلولی قرار گیرد. در این پژوهش، یک جت پلاسما آرگون با جریان متناوب در محدوده کیلوهرتز معرفی شده است که امکان تنظیم مستقل ولتاژ (۱ تا ۲۰ کیلوولت) و فرکانس (۱۸ تا ۲۸ کیلوهرتز) را فراهم می‌سازد و به‌عنوان بستری نوآورانه برای غلبه بر مقاومت سلولی عمل می‌کند. نتایج نشان داد که تنظیم دقیق پارامترهای الکتریکی، امکان کنترل تحویل پراکسید هیدروژن (H_2O_2) خارج‌سلولی را فراهم می‌سازد؛ گونه‌ای واکنش‌پذیر با عمر طولانی که نقش کلیدی در القای سمیت سلولی دارد. در رده سلولی HT29 که دارای مقاومت دارویی در سرطان کولورکتال است، غلظت H_2O_2 در محیط کشت در محدوده ۲۹۱ تا ۳۷۱ میکرومولار تنظیم شد. این دوز با سمیت سلولی وابسته به دوز همبستگی خطی بسیار بالا نشان داد ($R^2 = 0.995, p < 0.001$). پارامترهای بهینه‌شده (۱۰.۵ کیلوولت، ۲۸ کیلوهرتز) موجب اختلال در تعادل اکسید و احیای سلولی شده و بقای سلول‌ها را پس از ۳ دقیقه تیمار به ۹.۲٪ ± ۳.۶٪ کاهش دادند. این رویکرد توانست مسیر آنتی‌اکسیدانی Nrf2/Srx را که در مقاومت به جت پلاسما هلیوم نقش دارد، به‌طور مؤثر دور بزند. یافته‌های این مطالعه نشان می‌دهد که کنترل دقیق تحویل H_2O_2 از طریق جت پلاسما آرگون قابل تنظیم، راهبردی مؤثر برای مقابله با مقاومت ذاتی شیمی‌درمانی در سرطان کولورکتال بوده و این فناوری را به‌عنوان روشی نویدبخش در انکولوژی دقیق معرفی می‌نماید.

واژه‌های کلیدی: جت پلاسمای آرگون؛ سرطان کولورکتال؛ مقاومت به شیمی درمانی؛ پراکسید هیدروژن (H_2O_2)

Fusion Reactivity of Plasma with Anisotropic Lorentzian Distribution

F. Khoshdoun[1], M. Shahmansouri[2*]

[1]Department of Physics, Faculty of Science, Arak University, Arak, Islamic Republic of Iran
[2]Department of Atomic and Molecular Physics, Faculty of Physics, Alzahra University, Tehran, Islamic Republic of Iran
**Email: m.shahmansouri@alzahra.ac.ir*

تحلیل واکنش‌پذیری گداخت در پلاسما با توزیع لورنتزی ناهمسانگرد

فهیمه خوشدون[۱]، مهران شاه منصوری[۲*]

[۱]گروه فیزیک، دانشکده علوم، دانشگاه اراک، اراک، جمهوری اسلامی ایران
[۲]گروه فیزیک اتمی و مولکولی، دانشکده فیزیک، دانشگاه الزهرا، تهران، جمهوری اسلامی ایران

چکیده

توزیع‌های ناهمسانگرد و انحراف از تعادل سرعت نقشی اساسی در فیزیک پلاسما و فرایندهای همجوشی هسته‌ای ایفا می‌کنند. ظهور دُم‌های پرانرژی در توزیع‌های غیرتعادلی، جمعیت ذرات پرانرژی را افزایش داده و به‌دنبال آن، احتمال تونل‌زنی کوانتومی و در نتیجه نرخ واکنش‌های همجوشی را تقویت می‌کند. در این پژوهش، به بررسی اثر ناهمسانگردی در فضای سرعت و انحراف از تعادل ترمودینامیکی بر بهینه‌سازی بازده همجوشی می‌پردازیم. به‌طور خاص، مدل‌های توزیع غیرماکسولی، شامل توزیع‌های کاپا و کاپای ناهمسانگرد، تحلیل شده‌اند تا تأثیر آن‌ها بر واکنش‌پذیری همجوشی ارزیابی شود. نتایج ما نشان می‌دهند که توزیع‌های ناهمسانگرد در دماهای پایین عملکرد بهتری نسبت به توزیع‌های همسانگرد دارند، در حالی که در دماهای بالا، توزیع‌های همسانگرد غالب می‌شوند. این یافته‌ها بینش‌های جدیدی برای طراحی سامانه‌های همجوشی فراهم کرده و می‌توانند به بهبود کارایی فرایندهای همجوشی کمک کنند.

واژه‌های کلیدی: توزیع‌های غیرماکسولی؛ توزیع کاپا؛ نرخ واکنش‌های همجوشی؛ ناهمسانگردی سرعت؛ محصورسازی مغناطیسی

JOURNAL OF SCIENCES

ISLAMIC REPUBLIC OF IRAN

SUBSCRIPTION FORM

Please enter my annual subscription to the Journal of Sciences, Islamic Republic of Iran including 4 quarterly issues for the Year …… Vol. …… No. ……

	Iran	Other Countries
☐ Personal	R. 30,000	$ 80.00
☐ Institutional	R. 40,000	$ 100.00
☐ Student	R. 16,000	$ 50.00

☐ Check enclosed ☐ Bill me

Name:
Mailing Address: **City:** **Country:**

Check or money order must be made to order of:
Journal of Sciences, Islamic Republic of Iran, University of Tehran
Enghelab Ave., P.O. Box 13145-478, Tehran, Iran
Payment can be made via our transfer account.
Iran: Account No. 5225458914 Mellat Bank, Branch of University of Tehran, Central Income of Vice President of Research.
Foreign: Account No. 162454986 tejarat Bank, Branch of University of Tehran, Islamic Republic of Iran.
* Please allow 6-8 weeks for delivery.

JOURNAL OF SCIENCES

ISLAMIC REPUBLIC OF IRAN

SUBSCRIPTION FORM

Please enter my annual subscription to the Journal of Sciences, Islamic Republic of Iran including 4 quarterly issues for the Year …… Vol. …… No. ……

	Iran	Other Countries
☐ Personal	R. 30,000	$ 80.00
☐ Institutional	R. 40,000	$ 100.00
☐ Student	R. 16,000	$ 50.00

☐ Check enclosed ☐ Bill me

Name:
Mailing Address: **City:** **Country:**

Check or money order must be made to order of:
Journal of Sciences, Islamic Republic of Iran, University of Tehran
Enghelab Ave., P.O. Box 13145-478, Tehran, Iran
Payment can be made via our transfer account.
Iran: Account No. 5225458914 Mellat Bank, Branch of University of Tehran, Central Income of Vice President of Research.
Foreign: Account No. 162454986 tejarat Bank, Branch of University of Tehran, Islamic Republic of Iran.
* Please allow 6-8 weeks for delivery.

To:
Journal of Sciences
Islamic Republic of Iran

University of Tehran
Enghelab Ave., Tehran
P.O. Box 13145-478
Islamic Republic of Iran

Stamp

گیرنده:
تهران - مجله بین المللی علوم پایه جمهوری اسلامی ایران
تهران، خیابان انقلاب، دانشگاه تهران،
صندوق پستی ٤٧٨ - ١٣١٤٥

تمبر

Cover Legend:

"Abu Reyhan al-Biruni (973-1048) was one of the greatest scholars and philosopher in the history of Islamic civilization. He was equally well versed in a number of fields including mathematics, astronomy, physical and natural sciences, geography, history, chronology, and linguistics. George Sarton has called the first half of the 11th century «the age of Biruni». He wrote over 150 works, including 18 volumes on astronomy, 15 volumes on geography, 13 on literature, 12 volumes on astrology, 6 on religious ideology, 6 on astrolabes, 5 on timing devices, 5 on comets and 4 on the physics of light and optics. His scientific contributions also include the accurate determination of the density of eighteen different metals and minerals. He calculated the sine and cosine of angles from zero to 90 degrees and recorded them in a table. He made a remarkably accurate determination of the radius of the earth, which is very close to the latest estimate. He wrote dissertations regarding comets and meteorites, sun rays, light, twilight and dawn, time and space, astrolabes, methods to determine Qibla, the values of numerals, the relativity between precious stones and metals, ways of determining the longitude and latitude of cities, and the distances between them, and devised a method of measuring the earth's circumference, and complex problems of trigonometry. In addition to his native language (Persian) he also mastered Arabic, Sanskrit, Hebrew, Greek, Syriac and Turkish."

"Ref.: Noori-Daloii, M.R., J. Sci. I.R. Iran, Vol. 1, No. 1, pp. 2-3,. Autumn 1988"

JOURNAL OF SCIENCES

ISLAMIC REPUBLIC OF IRAN

University of Tehran

Indexed by Thomson Reuters, Scopus, Islamic World Science Citation Center (ISC), Scientific Information Database (SID) and/Abstracted in

Biological Abstracts
Biosis Selective Coverage Unique
Biosis Previews
Zoological Records
Chemical Abstracts
Methods in Organic Synthesis
Mathematical Reviews
Current Mathematical Publications
MathSciNet