

Kernel-Truncated Spline: Estimator Fleksibel untuk Regresi Nonparametrik

Rahmat Hidayat, Isma Muthahharah, dan Resmawan Resmawan



Volume 13, Issue 2, Pages 271–277, Aug. 2025

Diterima 28 Mei 2025, Direvisi 28 Juli 2025, Disetujui 6 Agustus 2025, Diterbitkan 8 Agustus 2025

To Cite this Article : R. Hidayat, I. Muthahharah, dan R. Resmawan, “Kernel-Truncated Spline: Estimator Fleksibel untuk Regresi Nonparametrik”, *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 13, no. 2, pp. 271–277, 2025, <https://doi.org/10.37905/euler.v13i2.33062>

© 2025 by author(s)

JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI



	Homepage	:	http://ejournal.ung.ac.id/index.php/euler/index
	Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
	Frequency	:	Three times a year
	Publication Language	:	English (preferable), Indonesia
	DOI	:	https://doi.org/10.37905/euler
	Online ISSN	:	2776-3706
	License	:	Creative Commons Attribution-NonCommercial 4.0 International License
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	http://ejournal.ung.ac.id/index.php/euler/oai
	Google Scholar ID	:	QF_r-gAAAAJ
	Email	:	euler@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



Jambura Journal of Biomathematics



Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics

Kernel-Truncated Spline: Estimator Fleksibel untuk Regresi Nonparametrik

Rahmat Hidayat^{1,*}, Isma Muthahharah¹, Resmawan Resmawan²

¹Program Studi Statistika, Universitas Negeri Makassar, Makassar 90224, Indonesia

²Program Studi Matematika, Universitas Negeri Gorontalo, Bone Bolango 96554, Indonesia

ARTICLE HISTORY

Diterima 28 Mei 2025

Direvisi 28 Juli 2025

Disetujui 6 Agustus 2025

Diterbitkan 8 Agustus 2025

KATA KUNCI

Nonparametrik

Spline

Kernel

GCV

Regresi

KEYWORDS

Nonparametric

Spline

Kernel

GCV

Regression

ABSTRAK. Penelitian ini bertujuan untuk mengembangkan model regresi nonparametrik multivariabel menggunakan integrasi estimator Spline dengan pendekatan kernel. Pendekatan ini diusulkan sebagai solusi atas keterbatasan model nonparametrik yang umumnya menggunakan satu jenis estimator untuk seluruh variabel prediktor, tanpa memperhatikan pola masing-masing variabel. Dalam penelitian ini, variabel dengan karakteristik osilatif dimodelkan menggunakan Spline, sedangkan variabel dengan pola non-linier kompleks dimodelkan dengan estimator kernel Gaussian. Model dikonstruksi dan diestimasi menggunakan metode Ordinary Least Square (OLS), serta diterapkan pada data rata-rata lama sekolah di Provinsi Sulawesi Selatan. Hasil menunjukkan bahwa kombinasi dua titik knot pada komponen Spline dan bandwidth optimal pada komponen kernel menghasilkan nilai Generalized Cross Validation (GCV) minimum sebesar 0,142, lebih baik dibandingkan model dengan satu atau tiga titik knot. Model terbaik memiliki nilai koefisien determinasi (R^2) sebesar 91,214% dan Mean Squared Error (MSE) sebesar 0,0461. Temuan ini mengindikasikan bahwa pendekatan regresi campuran mampu memodelkan data dengan lebih fleksibel dan akurat, khususnya dalam konteks data sosial multivariabel.

ABSTRACT. This study aims to develop a multivariable nonparametric regression model using a hybrid approach that combines Spline and kernel estimators. This method is proposed to address the limitations of conventional nonparametric models that typically apply a single type of estimator across all predictor variables, regardless of their individual patterns. In this approach, predictors with oscillatory patterns are modeled using truncated Spline regression, while variables exhibiting complex nonlinear behavior are modeled using a Gaussian kernel estimator. The combined model is constructed and estimated using the Ordinary Least Square (OLS) method and applied to data on the average years of schooling in South Sulawesi Province, Indonesia. Results indicate that the model using two Spline knots and an optimal bandwidth for the kernel component yields the lowest Generalized Cross Validation (GCV) value of 0.142, outperforming models with one or three knots. The best-fitting model achieves a coefficient of determination (R^2) of 91.214% and a Mean Squared Error (MSE) of 0.0461. These findings suggest that the hybrid regression approach offers greater flexibility and accuracy in modeling multivariable social data.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. **Editorial of EULER:** Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Pendahuluan

Berbagai jenis estimator telah dikembangkan dalam kerangka regresi nonparametrik, diantaranya estimator kernel, spline, polinomial lokal, wavelet, dan deret Fourier. Estimator kernel telah menjadi fokus sejumlah studi, seperti yang dibahas pada [1, 2]. Estimator spline juga banyak diteliti dalam berbagai bidang terapan, sebagaimana ditunjukkan pada [3, 4]. Sementara itu, beberapa pendekatan lain juga telah dilakukan dalam berbagai penelitian sejenis seperti pendekatan polinomial lokal [5], wavelet [6], dan deret Fourier [7].

Estimator kernel merupakan salah satu pendekatan awal yang secara luas digunakan dalam regresi nonparametrik. Seiring perkembangannya, pendekatan ini telah diaplikasikan dalam berbagai konteks. Sebagai contoh, Lee dkk. [8] mengem-

bangkan metode *kernel smoothing* untuk regresi multinomial, sedangkan Eisen dan Ribeire [9] menurunkan distribusi asimtotik dari estimator kernel berbobot untuk data longitudinal. Studi lain yang dilakukan oleh Jiang dkk. [10] menerapkan regresi kernel dalam menganalisis ketergantungan internet terhadap durasi penggunaan di kalangan siswa sekolah menengah. Sementara itu, Chen dkk. [11] memodifikasi estimator kernel adaptif Nadaraya-Watson dalam konteks regresi nonparametrik.

Selain estimator kernel, pendekatan spline juga banyak digunakan dan telah mengalami penyempurnaan sejak dikembangkan oleh Reinsch pada tahun 1967 melalui formulasi masalah optimasi [8]. Estimator spline biasanya dikonstruksi dari basis fungsi seperti spline truncated dan B-spline. Spline truncated memungkinkan adanya perubahan pola kurva pada berbagai interval, sehingga unggul dalam menangkap karakteristik lokal data. Studi yang dilakukan oleh Adiguzel dan Cengiz [12] menunjukk-

*Penulis Korespondensi.

an bahwa spline truncated memberikan performa lebih baik dibandingkan dengan *Multivariate Adaptive Regression Spline* (MARS), khususnya untuk fungsi dengan satu variabel prediktor. Pendekatan ini juga dimanfaatkan oleh Hidayat dkk. [13, 14] untuk memodelkan kasus pengangguran, dan dalam kasus lain menunjukkan bahwa nilai *Mean Squared Error* (MSE) dari spline truncated secara konsisten lebih kecil dibandingkan regresi linier, menegaskan keunggulannya dalam mengakomodasi variasi fungsi. Di antara berbagai pendekatan nonparametrik tersebut, spline truncated menonjol karena keunggulannya dalam hal interpretasi statistik dan visualisasi model. Estimator ini dikenal fleksibel serta mampu menangkap perubahan pola lokal dalam data multivariat. Di sisi lain, estimator kernel memiliki keunggulan dalam memodelkan data dengan pola yang tidak diketahui, dengan tingkat konvergensi yang relatif cepat.

Namun demikian, sebagian besar model regresi nonparametrik multivariat yang dikembangkan selama ini mengasumsikan dua hal utama yang bersifat membatasi: (1) semua variabel prediktor dianggap memiliki pola hubungan yang seragam terhadap respon, dan (2) hanya satu jenis estimator diterapkan untuk seluruh variabel prediktor. Dalam praktiknya, pola masing-masing variabel prediktor sering kali berbeda. Oleh karena itu, penggunaan satu bentuk estimator secara seragam dapat menghasilkan model yang kurang sesuai dengan karakteristik data, sehingga meningkatkan nilai galat. Untuk mengatasi hal tersebut, dibutuhkan pendekatan yang lebih adaptif, seperti kombinasi antara estimator kernel dan spline truncated dalam regresi nonparametrik multivariat. Pendekatan serupa juga dilakukan oleh Ronning [15] yang menggabungkan estimator Fourier dan spline truncated menggunakan metode *Penalized Least Squares* (PLS).

Salah satu indikator penting dalam mengevaluasi kualitas sumber daya manusia (SDM) di suatu wilayah adalah *rata-rata lama sekolah* (RLS), yakni jumlah tahun rata-rata penduduk berusia 25 tahun ke atas menerima pendidikan formal. Semakin tinggi RLS, semakin besar pula kemungkinan penduduk memiliki keterampilan dan pengetahuan yang menunjang peningkatan taraf hidup. Namun demikian, capaian RLS dipengaruhi oleh berbagai faktor, seperti kondisi sosial, ekonomi, dan ketersediaan infrastruktur pendidikan.

Faktor sosial yang relevan antara lain adalah rata-rata jumlah anggota keluarga. Berdasarkan teori alokasi sumber daya keluarga yang dikembangkan oleh Eisen dan Ribeiro [9], rumah tangga dengan jumlah anggota lebih banyak, khususnya anak-anak, cenderung memiliki alokasi biaya pendidikan per anak yang lebih rendah. Kondisi ini dapat menyebabkan anak tidak melanjutkan pendidikan ke jenjang yang lebih tinggi. Studi-studi di negara berkembang menunjukkan adanya korelasi negatif antara jumlah anak dalam rumah tangga dan tingkat pendidikan yang dicapai.

Selain itu, pengeluaran per kapita merupakan indikator ekonomi yang signifikan dalam memengaruhi keputusan keluarga untuk menyekolahkan anak. Menurut Muda dan Hutapea [16], keluarga dengan pengeluaran yang lebih tinggi memiliki kapasitas lebih besar untuk membiayai pendidikan, termasuk biaya penunjang seperti buku, seragam, transportasi, dan bimbingan belajar. Studi dari World Bank juga menegaskan bahwa pengeluaran rumah tangga merupakan salah satu determinan utama partisipasi sekolah, khususnya di kalangan masyarakat berpenghasilan

rendah. Dengan demikian, wilayah dengan pengeluaran per kapita yang tinggi cenderung memiliki RLS yang lebih tinggi karena pendidikan dipandang sebagai investasi jangka panjang.

Faktor penting lainnya adalah ketersediaan fasilitas pendidikan, seperti jumlah sekolah dan guru. Jumlah sekolah yang mencukupi dan tersebar secara merata memungkinkan akses pendidikan yang lebih mudah, tanpa hambatan geografis maupun finansial. Menurut laporan UNESCO, salah satu penyebab utama rendahnya partisipasi sekolah di daerah terpencil adalah terbatasnya infrastruktur pendidikan. Selain itu, jumlah guru yang memadai dan berkualitas berperan penting dalam menciptakan proses belajar mengajar yang efektif. Daerah dengan rasio murid-guru yang tinggi cenderung mengalami penurunan kualitas pembelajaran dan peningkatan angka putus sekolah, yang pada akhirnya menurunkan nilai RLS.

Keberadaan fasilitas sekolah di tingkat desa atau kelurahan juga menjadi indikator penting akses pendidikan dasar. Jika suatu desa tidak memiliki fasilitas sekolah, maka anak-anak harus menempuh jarak jauh untuk bersekolah, yang menjadi hambatan terutama di wilayah dengan kondisi geografis sulit. Studi oleh [17] menunjukkan bahwa kehadiran sekolah dalam desa berkorelasi positif dengan partisipasi pendidikan, khususnya pada jenjang dasar dan menengah pertama.

Secara keseluruhan, peningkatan RLS tidak dapat dipisahkan dari interaksi antara aspek sosial-ekonomi dan infrastruktur pendidikan. Untuk mewujudkan pemerataan pendidikan dan peningkatan kualitas SDM, diperlukan pendekatan kebijakan yang menyeluruh, tidak hanya pada aspek pembangunan fisik sekolah, tetapi juga pada pemberdayaan ekonomi keluarga dan perencanaan kependudukan. Kolaborasi antara pemerintah pusat, pemerintah daerah, dan masyarakat menjadi sangat penting untuk menjamin akses dan keberlanjutan pendidikan bagi setiap anak.

Berdasarkan kasus tersebut, penelitian ini bertujuan untuk mengembangkan model regresi nonparametrik multivariat yang lebih adaptif melalui integrasi estimator spline truncated dan kernel Gaussian, yang disesuaikan dengan karakteristik masing-masing variabel prediktor. Pemilihan metode ini diharapkan mampu meningkatkan fleksibilitas dan akurasi model dalam merepresentasikan hubungan kompleks antar variabel sosial. Kontribusi utama dari penelitian ini terletak pada usulan dan implementasi model regresi campuran yang memadukan dua pendekatan estimator nonparametrik secara selektif. Pendekatan ini memberikan alternatif metodologis yang lebih unggul dibandingkan metode konvensional yang hanya mengandalkan satu jenis estimator, serta berpotensi diterapkan secara luas dalam analisis data sosial multivariat yang kompleks.

2. Metode

Penelitian ini bertujuan untuk membentuk alternatif model dalam regresi nonparametrik. Untuk mencapai tujuan tersebut, metode *truncated spline* multivariat dan *kernel* pada regresi nonparametrik digunakan untuk menghasilkan estimator campuran. Tahapan-tahapan metodologis disusun sebagai berikut:

1. Membangun model regresi nonparametrik:

$$y_i = \sum_{p=1}^q f_p(x_{pi}) + g(t_i) + \varepsilon_i. \quad (1)$$

Dalam persamaan ini, fungsi f diasumsikan sebagai fungsi *spline*, sedangkan g adalah fungsi *kernel*. Variabel respons dinyatakan sebagai y , dengan x dan t sebagai variabel prediktor.

- Membentuk model dalam bentuk matriks berdasarkan pers. (1):

$$y = \mathcal{S}(\kappa)\beta + \mathcal{D}(\alpha)y + \varepsilon. \tag{2}$$

- Mengestimasi parameter β menggunakan metode Ordinary Least Squares (OLS) dengan menyelesaikan permasalahan optimasi sebagai berikut:

$$\min_{\alpha, \theta} \left\{ ((\mathbf{I} - \mathcal{C}(\alpha))\mathbf{y} - \mathcal{S}(\kappa)\beta)^T ((\mathbf{I} - \mathcal{C}(\alpha))\mathbf{y} - \mathcal{S}(\kappa)\beta) \right\}. \tag{3}$$

- Mengoptimalkan fungsi pada pers. (3) dengan pendekatan diferensiasi parsial terhadap parameter:

$$\frac{\partial Q(\beta)}{\partial \beta},$$

dengan:

$$Q(\beta) = ((\mathbf{I} - \mathcal{C}(\alpha))\mathbf{y} - \mathcal{S}(\kappa)\beta)^T ((\mathbf{I} - \mathcal{C}(\alpha))\mathbf{y} - \mathcal{S}(\kappa)\beta).$$

- Mengasumsikan turunan parsial sama dengan nol untuk memperoleh titik stasioner:

$$\frac{\partial Q(\beta)}{\partial \beta} = 0.$$

- Memperoleh taksiran hasil dari model regresi nonparametrik melalui integrasi metode spline multivariat dan kernel:

$$\hat{\mu}(x_i, t_i) = \sum_{p=1}^q \hat{f}_p(x_{pi}) + \hat{g}(t_i). \tag{4}$$

Setelah memperoleh estimator, model yang dihasilkan kemudian diterapkan pada data rata-rata lama sekolah di Provinsi Sulawesi Selatan untuk menilai performa estimasi yang dihasilkan.

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil penerapan model regresi nonparametrik multivariat berbasis *kernel-spline truncated*, serta pembahasan mengenai performa model berdasarkan nilai Generalized Cross Validation (GCV), koefisien determinasi (R^2), dan Mean Squared Error (MSE).

3.1. Model Regresi Nonparametrik Multivariat Berbasis Kernel-Spline Truncated

Untuk sejumlah variabel prediktor q , fungsi regresi didekati menggunakan model spline sebagai berikut:

$$f(x_i) = \sum_{j=0}^1 \vartheta_j x_i^j + \sum_{j=1}^r \varphi_j(x_i - \kappa_j)_+. \tag{5}$$

Berdasarkan pers. (5), diperoleh model dalam bentuk aljabar yang dinyatakan sebagai:

$$\sum_{p=1}^q f_p(x_{pi}) = \mathbf{X}_0 \vartheta + \mathbf{X}_1(\kappa_1)\varphi_1 + \mathbf{X}_2(\kappa_2)\varphi_2 + \dots + \mathbf{X}_q(\kappa_q)\varphi_q, \tag{6}$$

yang dapat dinyatakan dalam bentuk matriks:

$$\sum_{p=1}^q f_p(x_{pi}) = [\mathbf{X}_0 \quad \mathbf{X}_1(\kappa_1) \quad \dots \quad \mathbf{X}_q(\kappa_q)] \begin{bmatrix} \vartheta \\ \varphi_1 \\ \vdots \\ \varphi_q \end{bmatrix} = \mathcal{S}(\kappa)\beta. \tag{7}$$

dimana

$$\mathbf{Z}(\kappa) = [\mathbf{X}_0 \quad \mathbf{X}_1(\kappa_1) \quad \dots \quad \mathbf{X}_p(\kappa_p)] \quad \text{dan} \quad \beta = \begin{bmatrix} \vartheta \\ \varphi_1 \\ \vdots \\ \varphi_q \end{bmatrix}.$$

Selanjutnya, dari pers. (5) diberikan fungsi kernel untuk variabel dengan pola non-linier kompleks yang dinyatakan sebagai:

$$\hat{g}_\alpha(t) = n^{-1} \sum_{i=1}^n B_{\alpha i}(t)y_i, \tag{8}$$

dimana

$$B_{\alpha i}(t) = \frac{H_\alpha(t - t_i)}{n^{-1} \sum_{j=1}^n H_\alpha(t - t_j)}; \quad H_\alpha(t - t_i) = \frac{1}{\alpha} H\left(\frac{t - t_i}{\alpha}\right).$$

Persamaan kernel pada pers. (8) berlaku untuk $t = t_1, t_2, \dots, t_n$, sehingga diperoleh:

$$\begin{bmatrix} n^{-1}B_{\alpha 1}(t_1) & n^{-1}B_{\alpha 2}(t_1) & \dots & n^{-1}B_{\alpha n}(t_1) \\ n^{-1}B_{\alpha 1}(t_2) & n^{-1}B_{\alpha 2}(t_2) & \dots & n^{-1}B_{\alpha n}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}B_{\alpha 1}(t_n) & n^{-1}B_{\alpha 2}(t_n) & \dots & n^{-1}B_{\alpha n}(t_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \mathbf{C}(\alpha)\mathbf{y}. \tag{9}$$

Selanjutnya, dari pers. (5), diperoleh model campuran regresi Spline-kernel sebagai berikut:

$$\mathbf{y} = \sum_{p=1}^q \hat{f}_p(x_{pi}) + \hat{g}_\alpha(t) + \varepsilon = \mathcal{S}(\kappa)\beta + \mathcal{D}(\alpha)\mathbf{y} + \varepsilon. \tag{10}$$

Model pada pers. (10), diestimasi menggunakan metode *Ordinary Least Square* (OLS), dengan Kurva Spline multivariabel untuk $\sum_{p=1}^q f_p(x_{pi}) = \mathcal{S}(\kappa)\beta$, yang diberikan oleh persamaan berikut:

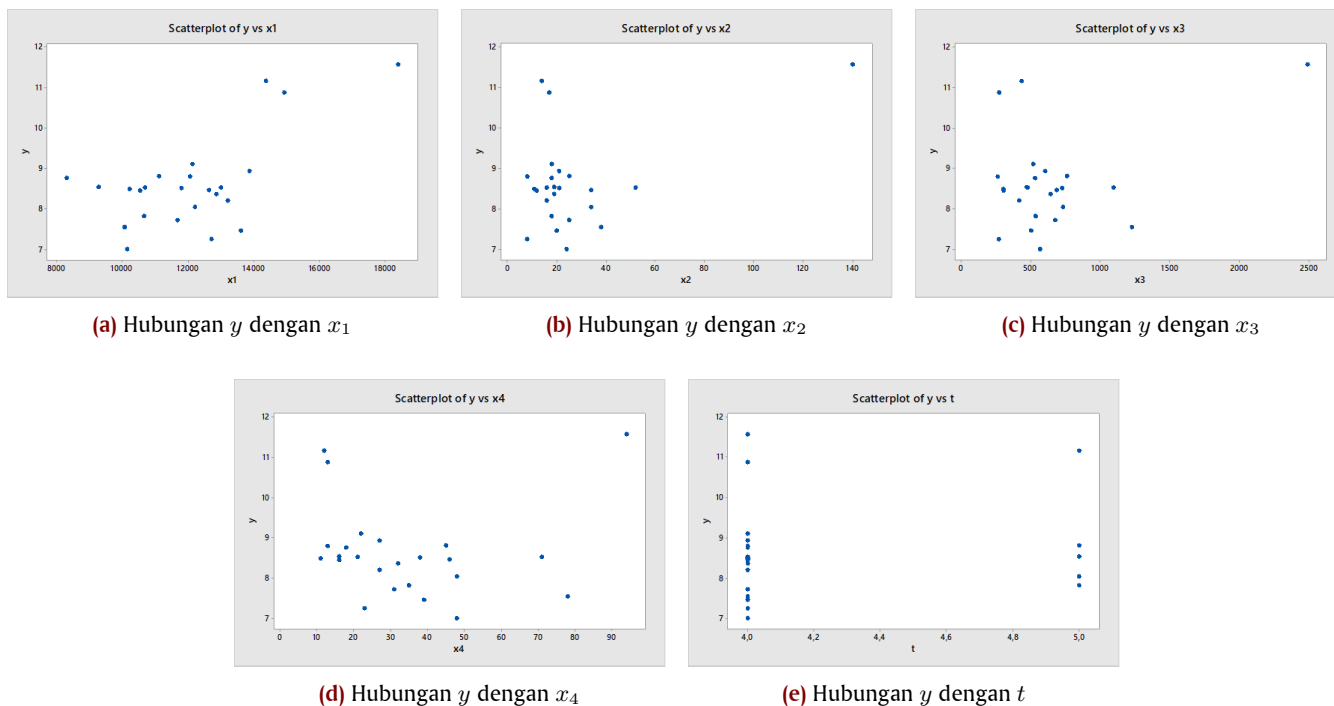
$$\sum_{p=1}^q \hat{f}_p(\alpha)(\mathbf{x}, t) = \mathcal{S}(\kappa)\hat{\beta}(\kappa, \alpha) = \mathcal{S}(\kappa) \left[(\mathcal{S}(\kappa)' \mathcal{S}(\kappa))^{-1} \mathcal{S}(\kappa)^T (\mathbf{I} - \mathbf{C}(\alpha)) \right] \mathbf{y}. \tag{11}$$

Bentuk ini menghasilkan persamaan:

$$\sum_{p=1}^q f_p(x_{pi}) = \mathbf{J}(\kappa, \alpha)\mathbf{y}, \tag{12}$$

dengan matrik \mathbf{J} :

$$\mathbf{J}(\kappa, \alpha) = \mathcal{S}(\kappa) \left[(\mathcal{S}(\kappa)^T \mathcal{S}(\kappa))^{-1} (\mathcal{B}(\kappa)^T) (\mathbf{I} - \mathbf{C}(\alpha)) \right].$$



Gambar 1. Scatter plot hubungan antar variabel

Berdasarkan pers. (9), bentuk estimator pada model regresi kernel dapat dinyatakan secara eksplisit, yaitu:

$$\hat{g}_{\alpha, \kappa}(t) = \mathbf{C}(\alpha) \mathbf{y}.$$

Dari proses tersebut, terbentuk estimator campuran yang mengintegrasikan regresi Spline dan kernel yang dinyatakan sebagai berikut:

$$\begin{aligned} \hat{\mu}(\mathbf{x}_i, t_i) &= \sum_{p=1}^q \hat{f}_p(x_{pi}) + \hat{g}(t_i) \\ &= \mathbf{J}(\kappa, \alpha) \mathbf{y} + \mathbf{C}(\alpha) \mathbf{y} \\ &= (\mathbf{J}(\kappa, \alpha) + \mathbf{C}(\alpha)) \mathbf{y} \\ &= \mathbf{B}(\kappa, \alpha) \mathbf{y}, \end{aligned} \tag{13}$$

dengan matriks \mathbf{B} :

$$\mathbf{B}(\kappa, \alpha) = \mathbf{J}(\kappa, \alpha) + \mathbf{C}(\alpha).$$

3.2. Visualisasi Pola Data dan Penentuan Estimator pada Data Riil

Untuk melihat pola hubungan antar variabel, dibuat visualisasi dalam bentuk scatter plot. Visualisasi hubungan antara variabel prediktor dan respon ditampilkan dalam scatter plot pada Gambar 1.

Scatter plot pada Gambar 1 menunjukkan bahwa hubungan antara variabel x_1 hingga x_4 terhadap respons y menunjukkan karakteristik yang sesuai untuk pendekatan model *spline*, yaitu pola sebaran data yang acak namun masih menunjukkan kecenderungan osilasi halus. Hal ini menunjukkan bahwa fungsi *spline* berpotensi menangkap variasi lokal dari data pada keempat variabel tersebut. Sebaliknya, untuk variabel x_5 (waktu t), pola osilasi yang mendasari tidak lagi terlihat secara konsisten, sehingga pendekatan dengan fungsi *spline* kurang sesuai. Oleh karena itu,

untuk variabel x_5 , pendekatan pemodelan yang lebih tepat adalah menggunakan fungsi *kernel* guna mengakomodasi bentuk pola yang tidak terstruktur secara jelas.

3.3. Nilai Generalized Cross-Validation (GCV) Berdasarkan Titik Knot

3.3.1. Titik Knot dan Bandwidth Optimal Pada Kasus Satu Titik Knot

Untuk mengoptimalkan pemilihan titik knot dan bandwidth, setiap variabel prediktor dalam komponen spline diberikan satu titik knot, sedangkan bandwidth ditentukan secara serempak. Pendekatan ini menghasilkan model regresi campuran kernel dan spline truncated multivariat sebagaimana ditunjukkan sebagai berikut:

$$\begin{aligned} \hat{y} &= \hat{\vartheta}_0^* + \hat{\vartheta}_{11}x_{1i} + \hat{\vartheta}_{12}x_{2i} + \hat{\vartheta}_{13}x_{3i} + \hat{\vartheta}_{14}x_{4i} + \hat{\varphi}_{11}(x_{1i} - \kappa_{11})_+ \\ &\quad + \hat{\varphi}_{12}(x_{2i} - \kappa_{12})_+ + \hat{\varphi}_{13}(x_{3i} - \kappa_{13})_+ + \hat{\varphi}_{14}(x_{4i} - \kappa_{14})_+ \\ &\quad + n^{-1} \sum_{i=1}^n \left[\frac{H_\alpha(t - t_i)}{n^{-1} \sum_{j=1}^n H_\alpha(t - t_j)} \right] y_i. \end{aligned}$$

Hasil penghitungan GCV dengan konfigurasi satu titik knot, disajikan pada Tabel 1.

Tabel 1. Nilai GCV dengan satu titik knot

Knot				Bandwidth	GCV
x_1	x_2	x_3	x_4	α	
8,468	25,13	352,46	30,092	0,512	0,289

Nilai *Generalized Cross Validation* (GCV) minimum sebesar 0,289 diperoleh ketika model menggunakan satu titik knot, sebagaimana ditunjukkan pada Tabel 1. Nilai GCV yang lebih rendah ini mengindikasikan bahwa model dengan satu titik knot memiliki kemampuan generalisasi yang lebih baik dibandingkan model

dengan jumlah knot yang lebih banyak. Dengan kata lain, model tersebut mampu meminimalkan galat prediksi terhadap data baru tanpa mengalami *overfitting*.

3.3.2. Titik Knot dan Bandwidth Optimal Pada Kasus Dua Titik Knot

Selanjutnya, pemodelan dilakukan dengan menggunakan dua titik knot pada masing-masing variabel prediktor, yang dirumuskan melalui persamaan berikut:

$$\hat{y} = \hat{\vartheta}_0^* + \hat{\vartheta}_{11}x_{1i} + \hat{\vartheta}_{12}x_{2i} + \hat{\vartheta}_{13}x_{3i} + \hat{\vartheta}_{14}x_{4i} + \hat{\varphi}_{11}(x_{1i} - \kappa_{11})_+ + \hat{\varphi}_{21}(x_{1i} - \kappa_{21})_+ + \hat{\varphi}_{12}(x_{2i} - \kappa_{12})_+ + \hat{\varphi}_{22}(x_{2i} - \kappa_{22})_+ + \hat{\varphi}_{13}(x_{3i} - \kappa_{13})_+ + \hat{\varphi}_{23}(x_{3i} - \kappa_{23})_+ + \hat{\varphi}_{14}(x_{4i} - \kappa_{14})_+ + \hat{\varphi}_{24}(x_{4i} - \kappa_{24})_+ + n^{-1} \sum_{i=1}^n \left[\frac{H_\alpha(t - t_i)}{n^{-1} \sum_{j=1}^n H_\alpha(t - t_j)} \right] y_i.$$

Model kemudian diuji dengan konfigurasi dua titik knot, dengan nilai GCV disajikan pada **Tabel 2**.

Tabel 2. Nilai GCV dengan dua titik knot

Knot				Bandwidth α	GCV
x_1	x_2	x_3	x_4		
8,111	8,387	356,922	20,376	1,337	0,142
12,184	19,832	619,457	50,557		

Berdasarkan **Tabel 2**, diketahui bahwa penggunaan dua titik knot menghasilkan nilai GCV minimum sebesar 0,142 pada salah satu kombinasi parameter. Nilai GCV yang lebih rendah ini menunjukkan bahwa model dengan dua titik knot memberikan tingkat kesesuaian yang lebih baik terhadap data dibandingkan model sebelumnya. Dengan demikian, konfigurasi dua knot dinilai lebih optimal dalam menangkap pola nonlinier pada variabel prediktor melalui pendekatan spline.

3.3.3. Titik Knot dan Bandwidth Optimal Pada Kasus Tiga Titik Knot

Tahapan selanjutnya dalam analisis adalah menerapkan pendekatan regresi *Kernel-Spline Truncated multivariat* dengan tiga titik knot. Model regresi yang dihasilkan dari konfigurasi tiga titik knot tersebut disajikan sebagai berikut:

$$\hat{y} = \hat{\vartheta}_0^* + \hat{\vartheta}_{11}x_{1i} + \hat{\vartheta}_{12}x_{2i} + \hat{\vartheta}_{13}x_{3i} + \hat{\vartheta}_{14}x_{4i} + \hat{\varphi}_{11}(x_{1i} - \kappa_{11})_+ + \hat{\varphi}_{21}(x_{1i} - \kappa_{21})_+ + \hat{\varphi}_{31}(x_{1i} - \kappa_{31})_+ + \hat{\varphi}_{12}(x_{2i} - \kappa_{12})_+ + \hat{\varphi}_{22}(x_{2i} - \kappa_{22})_+ + \hat{\varphi}_{32}(x_{2i} - \kappa_{32})_+ + \hat{\varphi}_{13}(x_{3i} - \kappa_{13})_+ + \hat{\varphi}_{23}(x_{3i} - \kappa_{23})_+ + \hat{\varphi}_{33}(x_{3i} - \kappa_{33})_+ + \hat{\varphi}_{14}(x_{4i} - \kappa_{14})_+ + \hat{\varphi}_{24}(x_{4i} - \kappa_{24})_+ + \hat{\varphi}_{34}(x_{4i} - \kappa_{34})_+ + n^{-1} \sum_{i=1}^n \left[\frac{H_\alpha(t - t_i)}{n^{-1} \sum_{j=1}^n H_\alpha(t - t_j)} \right] y_i.$$

Selanjutnya, hasil uji model dengan konfigurasi tiga titik knot menghasilkan nilai GCV yang disajikan pada **Tabel 3**.

Berdasarkan **Tabel 3**, terlihat bahwa kombinasi penggunaan tiga titik knot dengan jarak yang relatif merata memberikan nilai GCV minimum sebesar 0,588. Hal ini mengindikasikan bahwa model dengan tiga titik knot belum mampu memberikan tingkat presisi optimal sebagaimana yang diperoleh pada konfigurasi dengan satu atau dua titik knot.

Tabel 3. Nilai GCV dengan tiga titik knot

Knot				Bandwidth α	GCV
x_1	x_2	x_3	x_4		
8,254	9,556	434,753	12,874	0,812	0,8294
10,385	18,194	617,222	30,228		
15,812	24,654	795,754	31,192		

3.3.4. Perbandingan Nilai GCV Minimum

Untuk memberikan gambaran komparatif, nilai minimum GCV dari masing-masing konfigurasi satu, dua, dan tiga titik knot. Perbandingan ini bertujuan untuk mengidentifikasi struktur model terbaik dalam pendekatan regresi *Kernel-Spline Truncated multivariat* berdasarkan kriteria GCV terkecil. Selanjutnya, perbandingan nilai GCV minimum dari semua konfigurasi disajikan pada **Tabel 4**.

Tabel 4. Perbandingan nilai GCV minimum

Model	GCV
1 Knot	0,289
2 Knot	0,142
3 Knot	0,588

Perbandingan nilai GCV pada **Tabel 4** menunjukkan bahwa model Kernel dan *Spline Truncated multivariat* dengan dua titik knot menghasilkan nilai GCV minimum sebesar 0,142. Dengan demikian, model ini dipilih sebagai konfigurasi yang paling optimal untuk pemodelan Rata-rata Lama Sekolah di Provinsi Sulawesi Selatan. Nilai titik knot dan bandwidth yang diperoleh dari model ini akan digunakan dalam tahap pemodelan lanjutan sebagai parameter tetap untuk estimasi lebih lanjut.

Tabel 5. Hasil estimasi parameter

Variabel	Parameter	Estimasi
x_1	$\hat{\vartheta}_0^*$	-8,1931
	$\hat{\vartheta}_{11}$	0,0218
	$\hat{\varphi}_{11}$	-0,0221
	$\hat{\varphi}_{21}$	0,0163
x_2	$\hat{\vartheta}_{12}$	0,2181
	$\hat{\varphi}_{12}$	-0,4771
	$\hat{\varphi}_{22}$	0,2789
	$\hat{\vartheta}_{13}$	0,02377
x_3	$\hat{\varphi}_{13}$	-0,0276
	$\hat{\varphi}_{23}$	0,02311
	$\hat{\vartheta}_{14}$	-0,8621
x_4	$\hat{\varphi}_{14}$	1,4124
	$\hat{\varphi}_{24}$	-0,6192
bandwidth	α	1,337

3.4. Estimasi Parameter Model

Model regresi campuran Kernel dan *Spline Truncated multivariat* yang terbentuk berdasarkan konfigurasi optimal dua titik knot dapat dinyatakan sebagai berikut:

$$\hat{y} = \hat{\vartheta}_0^* + \hat{\vartheta}_{11}x_{1i} + \hat{\vartheta}_{12}x_{2i} + \hat{\vartheta}_{13}x_{3i} + \hat{\vartheta}_{14}x_{4i} + \hat{\varphi}_{11}(x_{1i} - \kappa_{11})_+ + \hat{\varphi}_{21}(x_{1i} - \kappa_{21})_+ + \hat{\varphi}_{12}(x_{2i} - \kappa_{12})_+ + \hat{\varphi}_{22}(x_{2i} - \kappa_{22})_+ + \hat{\varphi}_{13}(x_{3i} - \kappa_{13})_+ + \hat{\varphi}_{23}(x_{3i} - \kappa_{23})_+ + \hat{\varphi}_{14}(x_{4i} - \kappa_{14})_+$$

$$+ \hat{\varphi}_{24}(x_{4i} - \kappa_{24})_+ + n^{-1} \sum_{i=1}^n \left[\frac{H_\alpha(t - t_i)}{n^{-1} \sum_{j=1}^n H_\alpha(t - t_j)} \right] y_i.$$

Model regresi campuran spline dan kernel dengan dua titik knot menghasilkan estimasi parameter sebagaimana ditampilkan pada Tabel 5.

Dengan mensubstitusikan hasil estimasi parameter ke dalam model regresi campuran Kernel dan *Spline Truncated multivariat* dengan konfigurasi optimal dua titik knot, diperoleh persamaan sebagai berikut:

$$\begin{aligned} \hat{y} = & -8,1931 + 0,0218x_1 - 0,0221(x_1 - 8,111)_+ \\ & + 0,0163(x_1 - 12,184)_+ + 0,2181x_2 - 0,4771(x_2 - 8,387)_+ \\ & + 0,2789(x_2 - 19,832)_+ + 0,02377x_3 \\ & - 0,0276(x_3 - 356,922)_+ + 0,02311(x_3 - 619,457)_+ \\ & - 0,8621x_4 + 1,4124(x_4 - 20,376)_+ - 0,6192(x_4 - 50,557)_+ \\ & + \sum_{i=1}^{24} \left[\frac{0,0432 H\left(\frac{t-t_i}{1,337}\right)}{\sum_{j=1}^{24} 0,0432 H\left(\frac{t-t_j}{1,337}\right)} y_i \right]. \end{aligned}$$

3.5. Pembahasan

Model regresi campuran *Kernel* dan *Spline Truncated* multivariat dengan dua titik knot dan *bandwidth* optimum menghasilkan nilai koefisien determinasi R^2 sebesar 91,214% dengan nilai *Mean Squared Error* (MSE) sebesar 0,0461. Hal ini menunjukkan bahwa sebesar 91,214% variasi pada variabel *Rata-rata Lama Sekolah* dapat dijelaskan oleh lima variabel prediktor, yaitu:

- Rata-rata banyak anggota keluarga,
- Pengeluaran per kapita,
- Jumlah sekolah,
- Jumlah guru, dan
- Persentase Desa/Kelurahan yang memiliki fasilitas sekolah.

Model ini menunjukkan performa prediksi yang sangat baik, menandakan bahwa pendekatan nonparametrik campuran mampu menangkap hubungan nonlinier dan kompleks antara variabel prediktor dengan cukup akurat.

Namun, pendekatan ini memiliki sejumlah keterbatasan. Salah satu tantangan utama adalah proses pemilihan titik knot dan *bandwidth* yang optimal, yang memerlukan eksplorasi parameter secara menyeluruh serta komputasi yang cukup intensif, terutama saat jumlah variabel prediktor bertambah.

Selain itu, model ini belum secara eksplisit mengakomodasi interaksi antar variabel prediktor, dan efek heterogenitas spasial yang dapat berperan penting dalam fenomena sosial dan pendidikan. Padahal, dalam konteks sosial, hubungan antar variabel cenderung kompleks, tidak linier, dan tidak aditif. Oleh karena itu, pengembangan model lanjutan sangat disarankan, seperti dengan menggunakan pendekatan *Generalized Additive Models* (GAM), model regresi spasial nonparametrik, atau pendekatan *Bayesian nonparametrik*.

Temuan ini konsisten dengan studi terdahulu oleh Hidayat dkk. [18, 19] yang juga menggabungkan *smoothing spline* dan *kernel* dalam regresi nonparametrik dan melaporkan peningkatan akurasi model di atas 85%. Namun, berbeda dengan studi sebelumnya yang menerapkan estimator secara seragam, penelitian ini mengusulkan pendekatan pemisahan estimator berdasarkan

karakteristik masing-masing variabel. Hasilnya, nilai R^2 meningkat hingga 91,214%, mengindikasikan keunggulan dalam fleksibilitas dan akurasi model.

4. Kesimpulan

Penelitian ini mengusulkan dan menerapkan model regresi nonparametrik multivariat berbasis integrasi estimator *Spline* dan *Kernel*. Pendekatan ini dirancang untuk menangani berbagai pola hubungan nonlinier antara variabel prediktor dan respons, yang sering kali tidak dapat dijelaskan secara optimal oleh pendekatan tunggal. Melalui proses estimasi menggunakan metode *Ordinary Least Square* (OLS), diperoleh model terbaik dengan konfigurasi dua titik *knot* dan satu nilai *bandwidth* optimum. Model tersebut menghasilkan nilai *Generalized Cross Validation* (GCV) minimum sebesar 0,142, nilai koefisien determinasi R^2 sebesar 91,214%, dan *Mean Squared Error* (MSE) sebesar 0,0461. Hasil ini menunjukkan bahwa model mampu menjelaskan variabilitas *Rata-rata Lama Sekolah* (RLS) di Provinsi Sulawesi Selatan dengan tingkat akurasi yang sangat tinggi.

Model ini memperlihatkan bahwa pendekatan fleksibel dengan menggabungkan *Spline* dan *Kernel* dapat meningkatkan kualitas estimasi secara signifikan, khususnya dalam konteks data sosial yang kompleks. Variabel-variabel seperti rata-rata jumlah anggota keluarga, pengeluaran per kapita, jumlah sekolah, jumlah guru, serta ketersediaan fasilitas sekolah berhasil dimodelkan secara lebih realistis, sesuai dengan pola hubungan yang melekat pada masing-masing variabel.

Meskipun demikian, pengembangan lebih lanjut masih diperlukan. Model ini belum mengakomodasi kemungkinan adanya interaksi antar variabel prediktor maupun efek heterogenitas spasial yang mungkin relevan dalam konteks geografis dan sosial. Oleh karena itu, sebagai arah penelitian selanjutnya, model ini dapat diperluas dengan menambahkan komponen spasial atau interaksi prediktor guna menjangkau struktur hubungan yang lebih kompleks dan realistis.

Secara keseluruhan, model regresi campuran *Kernel* dan *Spline* multivariat memberikan kontribusi penting dalam metodologi analisis data nonparametrik, serta memiliki potensi besar untuk dikembangkan dan diimplementasikan dalam berbagai bidang ilmu terapan.

Kontribusi Penulis. Rahmat Hidayat: Konseptualisasi, metodologi, perangkat lunak, validasi, penulisan–persiapan draf asli. Isma Muthahharah: Konseptualisasi, investigasi, validasi, penulisan–tinjauan dan penyuntingan, supervisi. Resmawan: Metodologi, validasi, penulisan–tinjauan dan penyuntingan. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Ucapan Terima Kasih. Para penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penelitian ini dan dalam penyusunan manuskrip. Kami sangat menghargai editor dan reviewer atas masukan serta dukungannya dalam menyempurnakan karya ini.

Pembiayaan. Penelitian ini tidak menerima pendanaan eksternal.

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Ketersediaan Data. Tidak tersedia.

Referensi

- [1] C. Hazlett, "Kernel Balancing: A Flexible Non-Parametric Weighting Procedure for Estimating Causal Effects," *Stat. Sin.*, vol. 30, no. 3, pp. 1155–1189, 2020.
- [2] S. Węglarczyk, "Kernel density estimation and its application," *ITM Web Conf.*, vol. 23, p. 00037, 2018, doi: [10.1051/itmconf/20182300037](https://doi.org/10.1051/itmconf/20182300037).
- [3] A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, and M. Schmid, "A review of spline function procedures in R," *BMC Med. Res. Methodol.*, vol. 19, no. 1, p. 46, 2019, doi: [10.1186/s12874-019-0666-3](https://doi.org/10.1186/s12874-019-0666-3).
- [4] J. Gauthier, Q. V. Wu, and T. A. Gooley, "Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians," *Bone Marrow Transplant.*, vol. 55, pp. 675–680, 2020, doi: [10.1038/s41409-019-0679-x](https://doi.org/10.1038/s41409-019-0679-x).
- [5] J. P. Cohen, C. C. Coughlin, and J. M. Clapp, "Local Polynomial Regressions versus OLS for Generating Location Value Estimates," *The Journal of Real Estate Finance and Economics*, No. 2015-14, doi: [10.20955/wp.2015.014](https://doi.org/10.20955/wp.2015.014).
- [6] T. Guo *et al.*, "A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities," *IEEE Access*, vol. 10, pp. 58869–58903, 2022, doi: [10.1109/ACCESS.2022.3179517](https://doi.org/10.1109/ACCESS.2022.3179517).
- [7] W. Yao, Y. Weng, and J. M. Catchmark, "Improved cellulose X-ray diffraction analysis using Fourier series modeling," *Cellulose*, vol. 27, no. 10, pp. 5563–5579, 2020, doi: [10.1007/s10570-020-03177-8](https://doi.org/10.1007/s10570-020-03177-8).
- [8] C. Lee, D. Koo, and D. W. Zingg, "Comparison of B-Spline Surface and Free-Form Deformation Geometry Control for Aerodynamic Optimization," *AIAA J.*, vol. 55, no. 1, pp. 228–240, 2017, doi: [10.2514/1.j055102](https://doi.org/10.2514/1.j055102).
- [9] M. Eisen and A. Ribeiro, "Optimal Wireless Resource Allocation With Random Edge Graph Neural Networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 2977–2991, 2020, doi: [10.1109/TSP.2020.2988255](https://doi.org/10.1109/TSP.2020.2988255).
- [10] H. Jiang *et al.*, "Mobile Internet Quality Estimation using Self-Tuning Kernel Regression," *arXiv preprint*, Nov. 2023, doi: [10.48550/arXiv.2311.05641](https://doi.org/10.48550/arXiv.2311.05641).
- [11] B. Chen, J. Liang, N. Zheng, and J. C. Príncipe, "Kernel least mean square with adaptive kernel size," *Neurocomputing*, vol. 191, pp. 95–106, 2016, doi: [10.1016/j.neucom.2016.01.004](https://doi.org/10.1016/j.neucom.2016.01.004).
- [12] M. B. Adiguzel and M. A. Cengiz, "Model selection in multivariate adaptive regressions splines (MARS) using alternative information criteria," *Heliyon*, vol. 9, no. 9, 2023.
- [13] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, "The Regression Curve Estimation by Using Mixed Smoothing Spline and Kernel (MsS-K) Model," *Commun. Stat. Theory Methods*, vol. 50, no. 17, pp. 3942–3953, 2020, doi: [10.1080/03610926.2019.1710201](https://doi.org/10.1080/03610926.2019.1710201).
- [14] R. Hidayat *et al.*, "An Extended Model of Penalized Spline with The Addition of Kernel Functions in Nonparametric Regression Model," *Appl. Math. Inf. Sci.*, vol. 13, no. 3, pp. 453–460, 2019, doi: [10.18576/amis/130318](https://doi.org/10.18576/amis/130318).
- [15] F. Ronning, "The Role of Fourier Series in Mathematics and in Signal Theory," *Int. J. Res. Undergrad. Math. Educ.*, vol. 7, pp. 189–210, 2021, doi: [10.1007/s40753-021-00134-z](https://doi.org/10.1007/s40753-021-00134-z).
- [16] I. Muda and A. A. F. Hutapea, "Influence of capital expenditure and income original region to the income per capita in Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 126, p. 012065, 2018, doi: [10.1088/1755-1315/126/1/012065](https://doi.org/10.1088/1755-1315/126/1/012065).
- [17] X. Zhang and S. Rozelle, "Education Universalization, Rural School Participation, and Population Density," *China World Econ.*, vol. 30, no. 4, pp. 4–30, 2022, doi: [10.1111/cwe.12426](https://doi.org/10.1111/cwe.12426).
- [18] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, "A Reproducing Kernel Hilbert Space Approach and Smoothing Parameters Selection In Spline-Kernel Regression," *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 2, pp. 465–475, 2019.
- [19] R. Hidayat, M. Ilyas, and Y. Yuliani, "Spline model in the case of cervical cancer patient resilience," *Trends in Sciences*, vol. 20, no. 8, pp. 6170–6180, 2023.