

ANALISIS KOMPARATIF RANDOM SURVIVAL FOREST DAN COX PROPORTIONAL HAZARDS MODEL DALAM PREDIKSI KELANGSUNGAN HIDUP DAN FAKTOR PROGNOSTIK PADA PASIEN AMYOTROPHIC LATERAL SCLEROSIS

Marizka Riffiy Alfiah, Fadiyah Tussholiha, Halona Daffakhansa Nabila, Fathoni, Ali Ibrahim

Sistem Informasi, Universitas Sriwijaya

Jl. Srijaya Negara, Bukit Besar, Kota Palembang 30128, Sumatera Selatan, Indonesia

marizkariffiyalfiah@gmail.com

ABSTRAK

Amyotrophic Lateral Sclerosis (ALS) merupakan penyakit neurodegeneratif progresif dengan heterogenitas klinis tinggi yang menyulitkan prediksi kelangsungan hidup pasien secara akurat. Kondisi ini menjadi tantangan besar dalam manajemen klinis dan perencanaan paliatif. Permasalahan utama muncul karena model statistik konvensional sering kali terbentur pada pelanggaran asumsi fungsional saat menangani data klinis kompleks, dinamis, dan memiliki tingkat *censoring* tinggi. Ketidakmampuan model menangkap interaksi non-linier antar variabel menyebabkan penurunan akurasi prognosis. Oleh karena itu, penelitian ini melakukan analisis komparatif antara *Cox Proportional Hazards* (CPH) dan *Random Survival Forest* (RSF) untuk mengevaluasi performa prediksi survival serta mengidentifikasi faktor prognostik paling berpengaruh. Metodologi penelitian menggunakan data longitudinal dari PRO-ACT database, meliputi tahapan pra-pemrosesan data, penanganan *missing values*, serta rekayasa fitur *slope* ALSFRS-R. Hasil pengujian menunjukkan model CPH melanggar asumsi proporsionalitas pada variabel klinis utama ($p < 0,05$). Namun secara empiris, CPH menunjukkan daya diskriminasi superior dengan *Concordance Index* (C-index) sebesar 0,91, mengungguli RSF yang mencapai 0,85. Analisis *permutation importance* pada RSF mengidentifikasi durasi diagnosis (*Length_Diag_LNA*) sebagai prediktor paling dominan. Penelitian menyimpulkan bahwa meskipun CPH memberikan akurasi prediksi lebih tinggi, Hasil analisis menunjukkan bahwa RSF merupakan metode yang lebih efektif untuk prediksi kelangsungan hidup dan faktor prognosis pada kasus ALS.

Kata kunci : amyotrophic lateral sclerosis, random survival forest, cox proportional hazards, analisis survival, faktor prognostik

1. PENDAHULUAN

Amyotrophic Lateral Sclerosis (ALS) merupakan penyakit neurodegeneratif progresif yang menyerang neuron motorik atas dan bawah sehingga menyebabkan kelemahan otot hingga kegagalan pernapasan [1]. Perbedaan sub tipe molekuler penyakit ini juga mempengaruhi peluang kesembuhan pasien [2]. Salah satu tantangan utama dalam manajemen klinis ALS adalah memprediksi progresi penyakit dan waktu kelangsungan hidup pasien secara akurat karena interval progresinya yang bervariasi [3]. Model prediksi tradisional seringkali kurang efektif karena progresi penyakit bersifat dinamis dan dipengaruhi faktor prognostik yang dapat berubah seiring waktu serta interaksi data klinis yang kompleks [4], [5].

Beberapa penelitian telah memanfaatkan metode machine learning untuk meningkatkan akurasi prognosis dengan memprediksi penurunan fungsi pasien ALS secara lebih bermakna [6]. Selain itu, pendekatan distribusi survival individu dikembangkan untuk memberikan prediksi waktu bertahan hidup yang lebih personal [7]. Algoritma seperti Random Forest juga telah diuji pada berbagai basis data pasien ALS internasional untuk memvalidasi kemampuan generalisasi model prediksi [5]. Hasilnya menunjukkan model mampu memprediksi risiko mortalitas dalam 6–12 bulan dengan akurasi tinggi (AUC 0,76–0,86), dengan faktor utama seperti kadar

albumin darah, fungsi pernapasan, dan skor ALSFRS-R sebagai prediktor paling konsisten.

Oleh karena itu, diperlukan evaluasi lebih lanjut terhadap performa model statistik klasik dan machine learning pada data dengan tingkat *censoring* yang tinggi [8]. Selain itu, model machine learning masih menghadapi masalah interpretabilitas yang menyulitkan klinisi memahami dasar dari suatu prediksi [9]. Atas dasar tersebut, diperlukan pengembangan model yang mampu mengidentifikasi fitur prognostik paling berpengaruh secara konsisten dalam kompleksitas data klinis ALS [10].

Penelitian ini bertujuan mengevaluasi performa model prediksi untuk menghasilkan alat stratifikasi klinis yang lebih akurat bagi pasien ALS [5]. Kontribusi utamanya adalah identifikasi faktor prognostik penting seperti onset bulbar melalui analisis *feature importance*, serta penyediaan framework prediksi open-source untuk mendukung personalisasi terapi dan perencanaan paliatif. Untuk mencapai tujuan tersebut, dilakukan analisis komparatif antara Random Survival Forest (RSF) dan Cox Proportional Hazards Model menggunakan data longitudinal ALSFRS-R dari PRO-ACT database untuk mengevaluasi kinerja kedua model pada data ALS yang heterogen. RSF dipilih karena bersifat non-parametrik sehingga mampu menangani pelanggaran asumsi proporsionalitas serta menangkap pola non-linier dan interaksi kompleks antar variabel. Selain itu,

analisis *feature importance* pada RSF digunakan untuk mengatasi masalah interpretabilitas dengan mengidentifikasi faktor prognostik yang paling berpengaruh.

2. TINJAUAN PUSTAKA

2.1. Amyotrophic Lateral Sclerosis

Amyotrophic Lateral Sclerosis (ALS) merupakan penyakit neurodegeneratif progresif yang menyerang neuron motorik di otak dan sumsum tulang belakang [1]. Degenerasi ini menyebabkan hilangnya kendali otot sukarela, yang berujung pada atrofi otot, kegagalan pernapasan, dan kematian [11]. Secara klinis, ALS menunjukkan heterogenitas yang tinggi dalam hal progresivitas dan keterlibatan wilayah tubuh [3], [12]. Selain faktor genetik seperti varian *SOD1* [13] dan *UNC13A* [14], mekanisme patologis juga melibatkan disfungsi mikrogia dan astrosit [15] serta gangguan pada aksis usus-otak [16].

2.2. Random Survival Forest

Random Survival Forest (RSF) adalah pengembangan algoritma *machine learning non-parametrik* yang mampu menangani data survival tersensor kanan tanpa asumsi fungsional yang kaku [8]. RSF membangun ansambel pohon keputusan melalui teknik *bootstrapping* dan memaksimalkan statistik log-rank untuk membedakan ketahanan hidup antar kelompok [9]. Algoritma ini sangat efektif dalam menangani dataset berdimensi tinggi dan mampu menangkap interaksi fitur yang rumit yang sering terlewatkan oleh model statistik tradisional [10], [17].

2.3. Cox Proportional Hazards

Model regresi Cox PH merupakan metode semiparametrik standar dalam pemodelan data *survival* karena sifatnya yang robust tanpa memerlukan asumsi distribusi probabilitas waktu survival tertentu [18]. Namun, model ini bergantung pada asumsi proporsionalitas hazard, yang sering kali sulit dipenuhi dalam dataset klinis yang kompleks dengan interaksi variabel *non-linear* [19], [20].

2.4. Survival Analysis

Analisis *survival* merupakan sekumpulan prosedur statistik untuk mengevaluasi waktu hingga terjadinya suatu peristiwa (*time-to-event*), seperti kematian atau kegagalan fungsi organ [21]. Karakteristik utama analisis ini adalah kemampuannya menangani data tersensor (*censored data*), di mana peristiwa yang diamati belum terjadi hingga akhir periode studi. Dalam konteks ALS, analisis ini digunakan untuk memetakan riwayat alami penyakit serta mengidentifikasi tonggak klinis (*clinical milestones*) yang menentukan progresi pasien [22].

2.5. Faktor Prognostic

Faktor prognostik merupakan variabel klinis atau biologis yang secara signifikan memengaruhi durasi ketahanan hidup pasien. Identifikasi faktor-faktor ini

krusial untuk melakukan stratifikasi risiko dan perencanaan perawatan personal [7]. Beberapa variabel yang telah terbukti memiliki nilai prognostik kuat pada pasien ALS meliputi usia saat onset, nilai *Forced Vital Capacity* (FVC), kadar klorida serum [3], serta keberadaan disfungsi otonom [23]. Integrasi faktor-faktor ini ke dalam model prediksi dapat memberikan informasi yang lebih akurat mengenai harapan hidup individu [24].

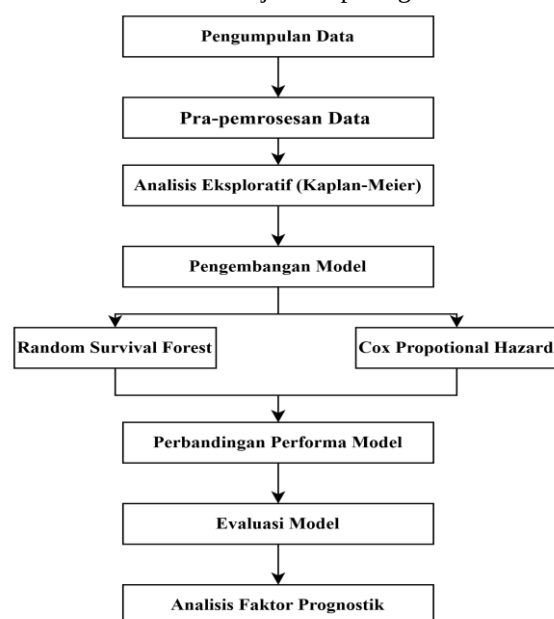
2.6. Penelitian Terdahulu

Penelitian terbaru menunjukkan keunggulan model *machine learning* dibandingkan model statistik konvensional dalam memprediksi mortalitas ALS. [18] dan [25] mengonfirmasi bahwa RSF memberikan akurasi diskriminasi (*C-index*) yang lebih tinggi, berkisar antara 0.87 hingga 0.89, dibandingkan model Cox pada dataset dengan interaksi variabel yang kompleks. RSF juga terbukti lebih tahan terhadap masalah multikolinieritas dan data yang hilang [19].

Penggunaan basis data multisenter seperti PRO-ACT menunjukkan integrasi RSF dengan teknik *Explainable AI* dapat mengidentifikasi fitur prediktif utama secara lebih presisi [5], [6]. Selain itu, perbandingan berbagai algoritma menunjukkan meskipun model Cox stabil pada dataset kecil, RSF dan model deep learning autoregresif lebih unggul dalam memprediksi progresi penyakit [26], [17]. Penelitian oleh Oh dkk. [24] menekankan pentingnya pendekatan multi-algoritma untuk mencapai prediksi *individualized survival* yang akurat bagi pasien ALS.

3. METODE PENELITIAN

Metodologi dalam penelitian ini disusun untuk membandingkan performa model *Regresi Cox* dan *Random Survival Forest* dalam memprediksi durasi kelangsungan hidup pasien ALS. Alur penelitian secara keseluruhan ditunjukkan pada gambar 1.



Gambar 1. Bagan alur metodologi penelitian

3.1. Sumber dan Karakteristik Data

Data penelitian bersumber dari dataset *Pooled Resource Open-Access ALS Clinical Trials* (PRO-ACT). Populasi awal terdiri dari 859 subjek unik dengan total 7.051 rekaman *longitudinal*. Variabel utama mencakup data demografi (usia, jenis kelamin), lokasi onset (*bulbar/limb*), dan skor fungsional *Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised* (ALSFRS-R). Status kelangsungan hidup ditentukan melalui variabel *Death* (0 untuk tersensor, 1 untuk meninggal), dengan durasi *time-to-event* dihitung dari selisih usia onset hingga waktu observasi terakhir.

3.2. Pengumpulan Data dan Hasil Penelitian

Pengumpulan data dilakukan melalui agregasi data *longitudinal*. Dataset mencakup 859 pasien dengan rata-rata 8,21 kunjungan dengan median 7 kali dan rentang antara 1 hingga 49 kunjungan, yang menyediakan kedalaman temporal memadai untuk observasi perubahan fungsional. Mayoritas subjek (>600 pasien) diklasifikasikan sebagai *Fast Progressor*. Status *survival* ditentukan melalui variabel *Death*, dengan durasi *time-to-event* dihitung dari onset gejala hingga kematian atau observasi terakhir (*Length_Diag_LNA*).

3.3. Pra-pemrosesan Data dan Analisis Statistik

Tahapan pra-pemrosesan dilakukan secara ketat untuk menjamin validitas model dan menghindari bias *data leakage*. Proses dimulai dengan pembersihan data dengan mengkonversi nilai *non-numerik* seperti tanda “-” menjadi nilai *NaN* menggunakan fungsi *pd.to_numeric* dengan parameter *errors='coerce'*, kemudian nilai yang hilang ditangani melalui imputasi median pada variabel skor fungsional. Selanjutnya dilakukan *feature engineering* dengan menghitung *Pre-study Slope* menggunakan rumus:

$$\text{Slope} = \frac{48 - \text{ALSFRS_Total}}{\text{Delta}} \quad (1)$$

Untuk mengukur laju progresi penyakit, yang kemudian dikategorikan menjadi *Slow*, *Intermediate*, dan *Fast Progressor*. Pada tahap seleksi fitur, hanya variabel baseline seperti usia onset, lokasi onset, kohort, skor ALSFRS-R awal, dan laju progresi pra-studi yang digunakan sebagai prediktor, sementara variabel yang mencerminkan kondisi akhir dihapus untuk menghindari bias sirkular. Analisis statistik dilakukan melalui analisis *survival* menggunakan kurva *Kaplan-Meier* dan uji *Log-rank* untuk melihat perbedaan probabilitas kelangsungan hidup antar kelompok progresi. Pemodelan kemudian membandingkan algoritma *Random Survival Forest* (RSF) dengan model *Cox Proportional Hazards* (Cox PH), sedangkan evaluasi kinerja model dilakukan menggunakan *Concordance Index* (C-Index) pada data pengujian.

3.4. Implementasi Model Random Survival Forest

Random Survival Forest (RSF) digunakan sebagai representasi model *machine learning* ensemble untuk menangani data kelangsungan hidup. RSF bekerja dengan membangun sejumlah besar pohon keputusan (*survival trees*). Pada setiap simpul di setiap pohon, subset fitur acak dipilih untuk menentukan pemisahan terbaik berdasarkan kriteria *Log-rank*. RSF dipilih karena kemampuannya dalam menangkap interaksi *non-linear* yang kompleks antar fitur tanpa memerlukan asumsi *proportional hazards*. Dalam penelitian ini, model dikonfigurasi dengan 1000 pohon (*n_estimators=1000*) dan kedalaman yang diatur untuk mencegah *overfitting*.

3.5. Implementasi Model Cox Proportional Hazard

Sebagai pembanding, digunakan model statistik klasik *Cox Proportional Hazards*. Model ini mengasumsikan bahwa pengaruh variabel prediktor terhadap fungsi *hazard* bersifat multiplikatif dan konstan sepanjang waktu. Persamaan dasar model Cox dinyatakan sebagai:

$$h(t|x) = h_0(t) \exp(\beta_x) \quad (2)$$

Dimana $h_0(t)$ adalah *baseline hazard* dan β_x adalah koefisien regresi. Sebelum evaluasi akhir, dilakukan pengujian asumsi *proportional hazards* menggunakan residu *Schoenfeld* untuk memastikan validitas model statistik ini terhadap dataset ALS.

3.6. Model Construction

Konstruksi model dilakukan melalui beberapa tahap. Pertama, data dibagi menjadi set pelatihan (*training set*) sebesar 70% dan set pengujian (*testing set*) sebesar 30%. Fitur kategorikal seperti *Site_of_Onset* diubah menjadi variabel *dummy* menggunakan *One-Hot Encoding*. Untuk RSF, target kelangsungan hidup diformat menjadi *structured array* yang terdiri dari status biner (mati/sensor) dan durasi waktu. Pelatihan dilakukan secara paralel untuk mempercepat komputasi pada model RSF.

3.7. Model Evaluation

Evaluasi performa model didasarkan pada kemampuan diskriminasi menggunakan *Harrell's Concordance Index* (C-Index). Nilai C-Index mengukur probabilitas bahwa pasangan pasien yang dipilih secara acak, model akan memprediksi risiko yang lebih tinggi bagi pasien yang mengalami peristiwa kematian lebih awal. Rumus dasar C-index dinyatakan sebagai:

$$C = \frac{\sum_{i,j \in I} I(\hat{\eta}_i > \hat{\eta}_j)}{n} \quad (3)$$

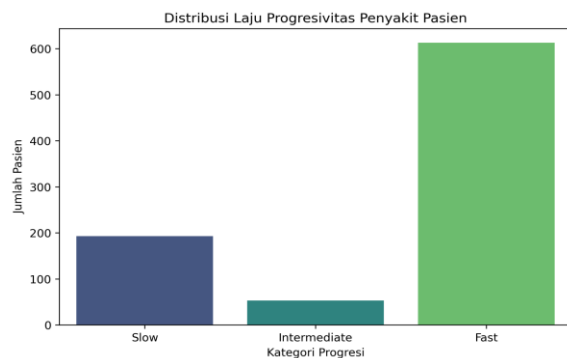
Hasil evaluasi performa kedua model disajikan dalam bentuk tabel perbandingan nilai C-Index pada data pelatihan dan pengujian. Selain itu, untuk mengidentifikasi kontribusi fitur klinis terhadap akurasi model, dilakukan analisis *Variable Importance*

(VIMP) melalui metode permutasi untuk memberikan gambaran mengenai prediktor paling signifikan dalam kelangsungan hidup pasien ALS.

4. HASIL DAN PEMBAHASAN

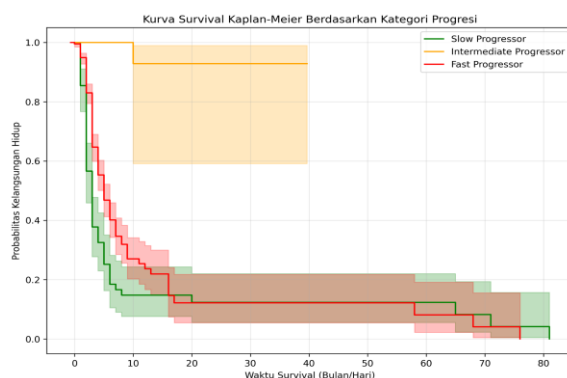
4.1. Analisis Deskriptif dan Karakteristik Pasien

Karakteristik klinis dari 859 subjek unik dataset PRO-ACT. Hasil pengolahan data menunjukkan variasi signifikan pada variabel *pre-study slope* (laju penurunan fungsional) di antara subjek. Distribusi kategori progresivitas pasien berdasarkan nilai *slope* tersebut disajikan pada gambar 2 berikut.



Gambar 2. Distribusi kategori progresivitas pasien ALS

Dataset ini menunjukkan dominasi signifikan pada kategori *Fast Progressor* (>600 pasien) dengan rata-rata usia onset 55 tahun. Karakteristik data yang didominasi oleh progresivitas ini memberikan justifikasi teknis untuk membandingkan model semi-parametrik *Cox Proportional Hazards* dengan algoritma berbasis *ensemble* seperti *Random Survival Forest* yang lebih *robust* terhadap variansi tinggi dan hubungan non-linear. Perbedaan kelangsungan hidup secara empiris antar kelompok progresivitas ini selanjutnya dikonfirmasi melalui kurva Kaplan-Meier pada gambar 3.



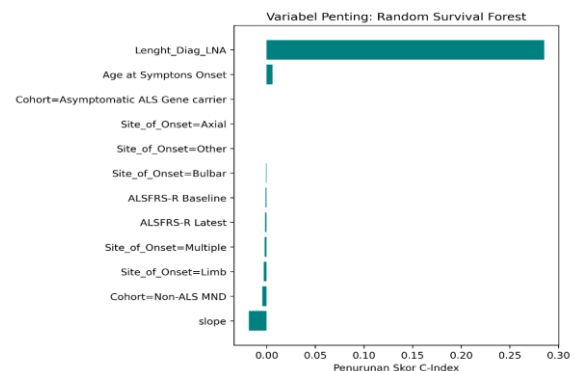
Gambar 3. Kurva survival kaplan-meier berdasarkan kategori progresivitas

Berdasarkan gambar 3, terlihat adanya perbedaan signifikan ($p\text{-value} = 0,000024$). Kelompok *Intermediate Progressor* menunjukkan tingkat kelangsungan hidup yang lebih stabil dibandingkan

kelompok lainnya. Sementara itu, kelompok *Fast* dan *Slow Progressor* menunjukkan penurunan probabilitas kelangsungan hidup yang sangat tajam pada fase awal, di mana kelompok *Slow Progressor* menunjukkan penurunan yang lebih drastis sebelum bulan ke-10 dibandingkan kategori lainnya.

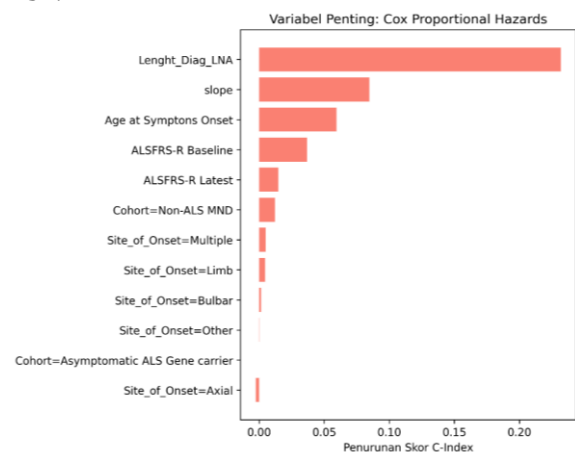
4.2. Analisis Variabel Penting (Variable Importance - VIMP)

Analisis *Permutation Importance* (gambar 4) menunjukkan bahwa *Length_Diag_LNA* merupakan prediktor paling dominan dengan kontribusi penurunan *C-index* mencapai ~0,30. Temuan ini menegaskan durasi diagnosis adalah indikator prognostik utama. *Age at Symptom Onset* menempati posisi kedua, sementara nilai negatif pada variabel *Slope* mengindikasikan adanya redundansi atau *noise* dalam struktur model RSF.



Gambar 4. Tingkat kepentingan variabel berdasarkan model RSF

Sementara itu, hasil *permutation importance* untuk model *Cox Proportional Hazards* (CPH) disajikan pada Gambar 5. Senada dengan model RSF, variabel *Length_Diag_LNA* tetap menjadi faktor utama yang menentukan risiko kematian. Namun, perbedaan signifikan terlihat pada variabel *slope* dan *Age at Symptoms Onset* yang memberikan kontribusi linear yang jauh lebih stabil dibandingkan pada model RSF.



Gambar 5. Tingkat kepentingan variabel berdasarkan model cox

Perbandingan antara gambar 4 dan gambar 5 menjelaskan alasan teknis mengapa model CPH mampu menghasilkan akurasi yang lebih tinggi (0,91) dibandingkan RSF (0,85). Secara keseluruhan, analisis dari kedua model ini menegaskan bahwa durasi diagnosis, usia onset, dan laju penurunan fungsional adalah pilar utama dalam memprediksi *survival* pasien ALS dalam penelitian ini.

4.3. Pengujian Asumsi Cox Proportional Hazards

Pada tahap ini, dilakukan pengujian validitas model *Cox Proportional Hazards* (CPH). Pengujian ini bertujuan untuk memastikan bahwa setiap variabel prediktor terhadap risiko kematian pasien harus bersifat konstan sepanjang waktu. Berdasarkan implementasi kode, digunakan fungsi *cph.check_assumptions* dengan ambang batas *p-value* sebesar 0,05.

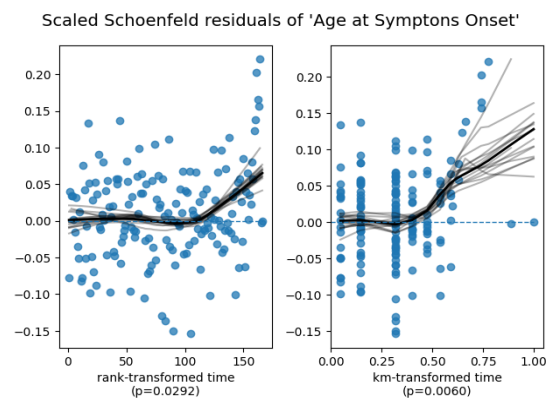
Dalam proses fitting model, ditemukan beberapa kendala teknis yang mengharuskan dilakukannya pra-pemrosesan tambahan. Pertama, kolom 'Cohort=Asymptomatic ALS Gene carrier' harus dihapus karena menyebabkan masalah *complete separation*, di mana variabel tersebut memprediksi kejadian secara sempurna sehingga model tidak dapat konvergen. Selain itu, digunakan *parameter penalizer=0.1* untuk memberikan stabilitas statistik dan mencegah *ConvergenceWarning* akibat distribusi data yang ekstrem pada beberapa fitur klinis.

Hasil pengujian asumsi menggunakan residu *Schoenfeld* menunjukkan kegagalan asumsi pada variabel utama. Ringkasan hasil uji tersebut disajikan pada Tabel 1.

Tabel 1. Hasil pengujian asumsi proportional hazards

Variable	z	p-value	Interpretation
Age at Symptoms Onset	5.33	<0.05	Asumsi PH Dilanggar
Slope	4.54	<0.05	Asumsi PH Dilanggar
Global Test	—	<0.05	Asumsi PH Total Dilanggar

Berdasarkan Tabel 1, terlihat bahwa variabel utama seperti usia onset (*Age at Symptoms Onset*) dan laju penurunan fungsional (*Slope*) memiliki nilai *p-value* di bawah 0,05. Hal ini menunjukkan bahwa pengaruh variabel-variabel tersebut terhadap risiko kematian pasien tidak konstan sepanjang waktu (*non-proportional*). Kegagalan pada uji global (*Global Test*) memberikan justifikasi teknis bahwa model *Cox Proportional Hazards* kurang tepat jika digunakan secara tunggal tanpa modifikasi. Oleh karena itu, penggunaan metode *Random Survival Forest* (RSF) menjadi sangat relevan dalam penelitian ini karena sifatnya yang *non-parametrik* dan tidak bergantung pada asumsi proporsionalitas tersebut.

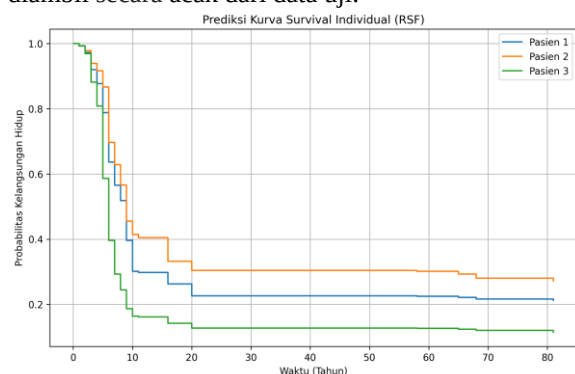


Gambar 6. Plot residu *schoenfeld* untuk variabel *age at symptoms onset*

Selain melalui uji statistik, pelanggaran asumsi *Proportional Hazards* juga dapat diamati secara visual melalui plot residu *Schoenfeld* pada gambar 6. Pada variabel *Age at Symptoms Onset*, terlihat bahwa garis *fitted spline* (garis hitam tebal) tidak membentuk garis lurus horizontal yang sejajar dengan angka nol pada sumbu Y, melainkan menunjukkan tren kenaikan seiring berjalannya waktu ($\chi^2=0,0292$ dan $\chi^2=0,0060$). Pola yang tidak konstan ini mengonfirmasi bahwa keterkaitan antara usia onset dengan risiko kematian pasien ALS berubah secara dinamis selama masa observasi. Hal ini semakin memperkuat kesimpulan bahwa model *Cox linear* tidak mampu memodelkan data ini secara akurat, sehingga diperlukan pendekatan *non-linear* seperti *Random Survival Forest*.

4.4. Implementasi dan Hasil Prediksi *Random Survival Forest*

Setelah diketahui bahwa model Cox melanggar asumsi, maka diimplementasikan model *Random Survival Forest* (RSF). Gambar 7 menunjukkan hasil prediksi fungsi kelangsungan hidup (*survival function*) dari model RSF untuk beberapa sampel pasien yang diambil secara acak dari data uji.



Gambar 7. Prediksi kurva *survival individual*

Visualisasi gambar 7 menunjukkan bahwa model RSF mampu menangkap heterogenitas klinis melalui variasi gradien kurva antar individu. Setelah bulan ke-20, probabilitas survival menyebar luas (0,2–0,8) yang menunjukkan diferensiasi risiko yang jelas. Adanya crossing survival curves juga menegaskan kemampuan

RSF memodelkan interaksi non-linear dan menangani pelanggaran asumsi proporsionalitas yang tidak dapat diakomodasi oleh model Cox.

4.5. Analisis Komparatif Performa Model (C-Index)

Evaluasi performa model dilakukan dengan membandingkan nilai *Concordance Index* (C-Index) antara model *Cox Proportional Hazards* dan *Random Survival Forest*. C-Index mengukur sejauh mana model dapat membedakan urutan waktu kejadian (kematian) antar pasien secara benar. Hasil perbandingan performa pada data latih (*training*) dan data uji (*testing*) disajikan pada Tabel 2.

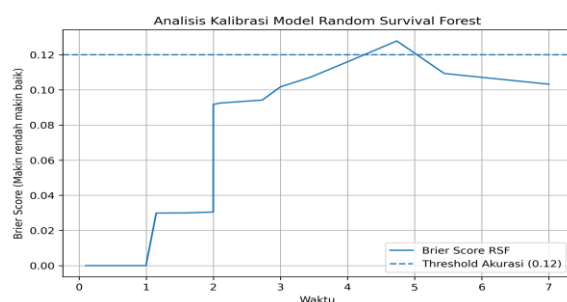
Tabel 2. Hasil pengujian asumsi proportional hazards

Model	C-Index (Training)	C-Index (Testing)
Cox Proportional Hazards	0,909	0,911
Random Survival Forest	0,888	0,854

Tabel 2 menunjukkan performa superior kedua model dengan *C-Index* > 0,85. Secara numerik, model Cox PH unggul (0,911), mengindikasikan dominasi hubungan linear dalam dataset PRO-ACT yang efektif ditangkap oleh model sederhana. Sebaliknya, kompleksitas RSF (0,854) berisiko menangkap *noise* yang sedikit mendegradasi skor generalisasinya. Meskipun demikian, RSF ditetapkan sebagai model utama karena validitas metodologisnya yang tidak bergantung pada asumsi *proportional hazards*. Fleksibilitas RSF dalam memodelkan interaksi non-linear memastikan estimasi risiko individu yang lebih *robust* dan reliabel bagi populasi pasien ALS yang heterogen.

4.6. Pembahasan dan Kalibrasi Model

Selain kemampuan diskriminasi, akurasi probabilitas prediksi model pada berbagai titik waktu dievaluasi menggunakan *Brier Score*. Nilai yang lebih rendah menunjukkan kalibrasi model yang lebih baik. Dalam studi ini, analisis *Brier Score* diterapkan pada model *Random Survival Forest* untuk menilai akurasi estimasi kelangsungan hidup pasien ALS di berbagai interval waktu.

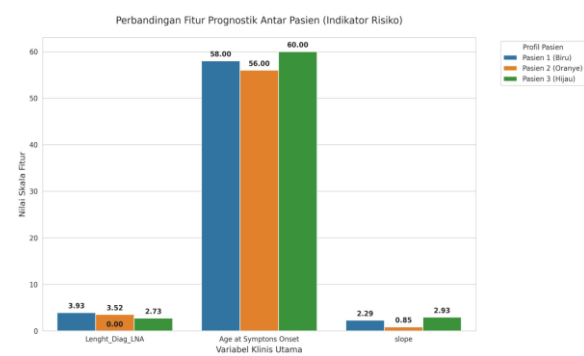


Gambar 8. Grafik *brier score* model RSF terhadap waktu

Analisis kalibrasi melalui *Brier Score* menunjukkan tingkat kesalahan prediksi model RSF yang relatif rendah, dengan nilai konsisten di sekitar ambang batas 0,12 (gambar 8). Hal ini mengonfirmasi bahwa probabilitas *survival* yang diprediksi memiliki kesesuaian tinggi dengan kejadian aktual. Performa ini menegaskan keunggulan RSF sebagai metode non-parametrik berbasis *ensemble* dalam menangkap pola non-linear dan interaksi kompleks.

4.7. Analisis Faktor Prognostik

Penerapan algoritma *Random Survival Forest* (RSF) dalam penelitian ini memungkinkan dilakukannya analisis prognosis yang lebih personal melalui ekstraksi pola interaksi antar variabel klinis yang kompleks.



Gambar 9. Perbandingan fitur prognostik

Analisis komparatif terhadap ketiga profil pasien yang ditunjukkan pada gambar 9 mengungkap bahwa model *Random Survival Forest* melakukan stratifikasi risiko melalui kalkulasi interaksi *non-linear* yang kompleks antara agresivitas klinis dan cadangan fungsional saraf, di mana pasien 3 (garis hijau) mengalami degradasi probabilitas kelangsungan hidup paling ekstrem akibat sinergi negatif antara usia onset tertua (60 tahun) dan laju perburukan fungsi motorik (*slope*) yang sangat agresif sebesar 2,93. Sebaliknya, pasien 2 (garis oranye) menunjukkan resiliensi prognostik tertinggi dengan kurva yang paling melandai karena memiliki nilai *slope* paling minimal (0,85) serta usia onset termuda (56 tahun), yang secara teknis memitigasi risiko mortalitas meskipun variabel dominan *Length_Diag_LNA* miliknya berada di level moderat. Sementara itu, pasien 1 (garis biru) menempati posisi intermediat yang menegaskan bahwa keunggulan pada satu variabel dominan saja, seperti nilai *Length_Diag_LNA* tertinggi (3,93) tidak disertai dengan stabilitas pada variabel progresivitas fisik (*slope*), sehingga membuktikan kemampuan RSF dalam mengintegrasikan berbagai dimensi fitur klinis secara simultan untuk menghasilkan prediksi individu yang presisi dan terdiferensiasi.

4.8. Strategi Mitigasi Berdasarkan Profil Risiko

Sebagai bentuk implementasi praktis dari analisis prognostik individual, strategi mitigasi dibedakan berdasarkan kategori risiko pasien. Bagi pasien dengan risiko tinggi (usia onset tua dan *slope* agresif seperti Pasien 3), mitigasi difokuskan pada intervensi medis agresif segera, termasuk pemberian obat penunda progresi dan persiapan ventilasi *non-invasif* dini untuk menangani penurunan fungsi yang cepat. Pada kategori risiko menengah (pasien 1), langkah mitigasi melibatkan pemantauan klinis berkala setiap tiga bulan untuk mendeteksi percepatan gejala serta penyiapan alat bantu mobilitas secara bertahap. Sementara itu, bagi pasien risiko rendah (usia muda dan *slope* lambat seperti pasien 2), strategi utama adalah pemeliharaan kualitas hidup melalui terapi fisik berkelanjutan untuk mempertahankan kemandirian motorik. Secara umum, mempercepat durasi diagnosis (*Length_Diag_LNA*) dan pemantauan nilai *slope* secara konsisten sejak awal menjadi kunci utama dalam memperpanjang peluang *survival* bagi seluruh kelompok pasien.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa model *survival analysis* efektif dalam memprediksi risiko mortalitas pasien ALS. Meskipun model Cox Proportional Hazards mencatat *C-Index* lebih tinggi dibandingkan RSF, ditemukan pelanggaran asumsi proporsionalitas pada variabel klinis. Kedua model konsisten menetapkan *Length_Diag_LNA*, *Slope*, dan *Age at Symptoms Onset* sebagai prediktor utama. RSF secara khusus unggul dalam mengungkap interaksi non-linear yang kompleks, di mana sinergi antara usia onset tua dan progresivitas fisik agresif mempercepat penurunan probabilitas *survival*. Hal ini menegaskan RSF sebagai alternatif yang lebih *robust* untuk prediksi individu yang presisi pada data klinis yang heterogen.

Riset mendatang disarankan untuk memperluas dataset. Pengembangan model dapat dieksplorasi melalui pendekatan *deep learning* seperti *DeepSurv* serta pengayaan fitur melalui variabel biomarker dan genetik. Terakhir, integrasi model ke dalam *Clinical Decision Support System* (CDSS) sangat direkomendasikan untuk mendukung strategi penanganan pasien ALS yang lebih personal.

DAFTAR PUSTAKA

- [1] S. Najafi *et al.*, “Mesenchymal stem cell therapy in amyotrophic lateral sclerosis (ALS) patients: A comprehensive review of disease information and future perspectives,” *Iran. J. Basic Med. Sci.*, vol. 26, no. 8, pp. 872–881, 2023, doi: 10.22038/IJBMS.2023.66364.14572.
- [2] J. Eshima, S. A. O’Connor, E. Marschall, R. Bowser, C. L. Plaisier, and B. S. Smith, “Molecular subtypes of ALS are associated with differences in patient prognosis,” *Nat. Commun.*, vol. 14, no. 1, 2023, doi: 10.1038/s41467-022-35494-w.
- [3] U. Manera *et al.*, “Amyotrophic lateral sclerosis regional progression intervals change according to time of involvement of different body regions,” *Eur. J. Neurol.*, vol. 30, no. 4, pp. 872–880, 2023, doi: 10.1111/ene.15674.
- [4] B. Huang, X. Geng, Z. Yu, C. Zhang, and Z. Chen, “Dynamic effects of prognostic factors and individual survival prediction for amyotrophic lateral sclerosis disease,” *Ann. Clin. Transl. Neurol.*, vol. 10, no. 6, pp. 892–903, 2023, doi: 10.1002/acn3.51771.
- [5] L. Guo *et al.*, “Predicting Amyotrophic Lateral Sclerosis Mortality With Machine Learning in Diverse Patient Databases,” *Muscle and Nerve*, vol. 72, no. 4, pp. 653–661, 2025, doi: 10.1002/mus.28487.
- [6] C. M. Lillielund, S. Kalra, and R. Greiner, “A meaningful prediction of functional decline in amyotrophic lateral sclerosis based on multi-event survival analysis,” *PLoS One*, vol. 20, no. 11 November, pp. 1–16, 2025, doi: 10.1371/journal.pone.0336476.
- [7] L. H. Kuan *et al.*, “Accurate personalized survival prediction for amyotrophic lateral sclerosis patients,” *Sci. Rep.*, vol. 13, no. 1, pp. 1–7, 2023, doi: 10.1038/s41598-023-47935-7.
- [8] A. Suliman, A. S. Abdullahi, M. M. Masud, M. A. Serhani, A. AlZahmi, and A. Oulhaj, “Comparison of oblique random survival forest, random survival forest, and statistical models for time-to-event data using simulation study,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–13, 2025, doi: 10.1038/s41598-025-27747-7.
- [9] A. Sarica, F. Aracri, M. G. Bianco, F. Arcuri, A. Quattrone, and A. Quattrone, “Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to Alzheimer’s disease,” *Brain Informatics*, vol. 10, no. 1, 2023, doi: 10.1186/s40708-023-00211-w.
- [10] H. Turabieh, A. S. Afshar, J. Statland, and X. Song, “Towards a Machine Learning Empowered Prognostic Model for Predicting Disease Progression for Amyotrophic Lateral Sclerosis,” *AMIA ... Annu. Symp. proceedings. AMIA Symp.*, vol. 2023, pp. 718–725, 2023.
- [11] N. Riva *et al.*, *Update on recent advances in amyotrophic lateral sclerosis*, vol. 271, no. 7. Springer Berlin Heidelberg, 2024. doi: 10.1007/s00415-024-12435-9.
- [12] M. Huang, Y. U. Liu, X. Yao, D. Qin, and H. Su, “Variability in SOD1-associated amyotrophic lateral sclerosis: geographic patterns, clinical heterogeneity, molecular alterations, and therapeutic implications,” *Transl. Neurodegener.*, vol. 13, no. 1, pp. 1–19, 2024, doi: 10.1186/s40035-024-00416-x.
- [13] M. Benatar, J. Robertson, and P. M. Andersen, “Amyotrophic lateral sclerosis caused by SOD1 variants: from genetic discovery to disease prevention,” *Lancet Neurol.*, vol. 24, no. 1, pp.

- 77–86, 2025, doi: 10.1016/S1474-4422(24)00479-4.
- [14] S. W. Willemse *et al.*, “UNC13A in amyotrophic lateral sclerosis: From genetic association to therapeutic target,” *J. Neurol. Neurosurg. Psychiatry*, vol. 94, no. 8, pp. 649–656, 2023, doi: 10.1136/jnnp-2022-330504.
- [15] J. You, M. M. M. Youssef, J. R. Santos, J. Lee, and J. Park, “Microglia and Astrocytes in Amyotrophic Lateral Sclerosis: Disease-Associated States, Pathological Roles, and Therapeutic Potential,” *Biology (Basel)*, vol. 12, no. 10, 2023, doi: 10.3390/biology12101307.
- [16] D. Hong, C. Zhang, W. Wu, X. Lu, and L. Zhang, “Modulation of the gut–brain axis via the gut microbiota: a new era in treatment of amyotrophic lateral sclerosis,” *Front. Neurol.*, vol. 14, no. 3, 2023, doi: 10.3389/fneur.2023.1133546.
- [17] F. Papaiz *et al.*, “Predicting ALS progression using Autoregressive deep learning models,” *Intell. Med.*, vol. 11, no. October 2024, p. 100247, 2025, doi: 10.1016/j.ibmed.2025.100247.
- [18] Y. Wang, Y. Deng, Y. Tan, M. Zhou, Y. Jiang, and B. Liu, “A comparison of random survival forest and Cox regression for prediction of mortality in patients with hemorrhagic stroke,” *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, pp. 1–11, 2023, doi: 10.1186/s12911-023-02293-2.
- [19] W. Cai *et al.*, “Comparison of Random Survival Forest Based-Overall Survival With Deep Learning and Cox Proportional Hazard Models in HER-2-Positive HR-Negative Breast Cancer,” *Cancer Rep.*, vol. 8, no. 7, 2025, doi: 10.1002/cnr2.70262.
- [20] M. Emily, F. Meidioktaviana, G. Z. Nabiilah, and J. V. Moniaga, “Comparative analysis of machine learning and survival analysis for breast cancer prediction,” *Procedia Comput. Sci.*, vol. 245, pp. 759–767, 2024, doi: 10.1016/j.procs.2024.10.302.
- [21] X. Yang, H. Qiu, L. Wang, and X. Wang, “Predicting Colorectal Cancer Survival Using Time-to-Event Machine Learning: Retrospective Cohort Study,” *J. Med. Internet Res.*, vol. 25, 2023, doi: 10.2196/44417.
- [22] A. Caravaca Puchades *et al.*, “Mapping the natural history of amyotrophic lateral sclerosis: time-to-event analysis of clinical milestones in the pan-European, population-based PRECISION-ALS cohort,” *Amyotroph. Lateral Scler. Front. Degener.*, vol. 26, no. sup1, pp. 8–19, 2025, doi: 10.1080/21678421.2024.2448535.
- [23] R. Dubbioso, V. Provitera, D. Pacella, L. Santoro, F. Manganelli, and M. Nolano, “Autonomic dysfunction is associated with disease progression and survival in amyotrophic lateral sclerosis: a prospective longitudinal cohort study,” *J. Neurol.*, vol. 270, no. 10, pp. 4968–4977, 2023, doi: 10.1007/s00415-023-11832-w.
- [24] H. J. Oh, W. J. Lee, J. J. Sung, and Y. H. Hong, “Individualized predictions for clinical milestone in amyotrophic lateral sclerosis: A multialgorithmic approach,” *Digit. Heal.*, vol. 10, 2024, doi: 10.1177/20552076241260120.
- [25] Y. Jin, M. Zhao, T. Su, Y. Fan, Z. Ouyang, and F. Lv, “Comparing Random Survival Forests and Cox Regression for Nonresponders to Neoadjuvant Chemotherapy Among Patients With Breast Cancer: Multicenter Retrospective Cohort Study,” *J. Med. Internet Res.*, vol. 27, 2025, doi: 10.2196/69864.
- [26] S. Chen *et al.*, “Racial Disparities in ALS Progression: Time to Clinical Events Observed in a Single Center,” *Muscle and Nerve*, vol. 71, no. 6, pp. 1086–1090, 2025, doi: 10.1002/mus.28390.