A Cosine Hazard Basis for Parkinson's Disease Progression

David Calhas
david.calhas@fri.uni-lj.si
Faculty of Computer and Information Science
University of Ljubljana
Slovenia

Žiga Črv
Faculty of Computer and Information Science
University of Ljubljana
Slovenia

Vida Groznik
Faculty of Computer and Information Science
University of Ljubljana
Slovenia

Dejan Georgiev
Department of Neurology
University Medical Centre Ljubljana
Slovenia

Aleksander Sadikov
Faculty of Computer and Information Science
University of Ljubljana
Slovenia

Abstract

Prediction of Parkinson's disease (PD) symptom progression is critical for timely clinical intervention, yet conventional survival models often rely on restrictive proportional hazards assumptions or computationally intensive numerical integration. To avoid this, we introduce a positive and patient-conditioned cosine basis hazard with fixed frequencies evaluated on standardized log time, for which the cumulative hazard has a closed form expression. We evaluate our approach on longitudinal motor assessments and demographic data from the Parkinson's Precision Medicine Initiative cohort across key Hoehn and Yahr stage transitions. Our results suggest prediction of PD progression benefits from patient specific hazard shapes.

Keywords

survival analysis, neural networks, neurodegenerative disease, progression

1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder, whose symptoms are summarised by the Hoehn and Yahr scale (HY) [10]. This scale distinguishes clinically meaningful changes in laterality, postural stability, and functional independence. Currently, one of the key questions in PD research is: *how long will the patient stay in the current HY state?* This answers how fast/slow is the patient progressing. An accurate answer impacts the patient's quality of life, because clinicians are able to prescribe timely medications that condition PD progression. To address this key question, machine learning has been widely used to predict time to event of the next HY stage. Largely, studies focus on disease diagnosis, fixed time score prediction, and data driven progression sub types [8, 9]. For time-to-event, efforts have been made in a Singapore movement disorder cohort to build Kaplan–Meier curve estimations of stage specific transitions [17]. Some studies have also used longitudinal scale level trajectories, in a Cox model, for progression to HY stage 3 [14], while others link all motor clinical items to progression from baseline stage 1 and 2 to ≥ 3 of the HY scale [15]. But none of these address censored HY prognosis and item level longitudinal input.

Besides the PD application gap, we also question if the hazard shape in survival analysis for PD progression is *proportional* [5]. Instead of individuals sharing the hazard shape, we hypothesize that PD progression is oscillatory, where the hazard function, $h(t)$, is described by a cosine basis (see Figure 1) and independent from each individual. Recent parametric work models positive, non-monotone, and oscillatory hazards with closed form hazard and cumulative-hazard functions [4]. In contrast, our approach uses neural network predicted amplitudes, which are optimized via gradient descent. The use of neural networks to generate coefficients of temporal basis functions has been previously done [7], but our work derives the explicit closed-form cumulative hazard for a positive multi-cosine hazard, including the nontrivial standardized log-time transformation. To test it, we experiment several hazard shape assumptions, in addition to the proportional hazards, on a common neural network, that processes both demographic and clinical history symptom data. In our *cosine basis assumption*, the model predicts the coefficients of each sinusoid that form the hazard function over time. The objective function is defined by

$$\mathcal{L} = -\delta \log\left(h(t)\right) + \Delta(t), \tag{1}$$

which was originally proposed in the SODEN paper [16]. $\Delta(t) = \int_t^\infty h(\tau)\mathrm{d}\tau$ is the cumulative hazard function, which can be directly derived from $h(t)$ and used to obtain the survival function, $S(t) = \exp(-H(t))$. In addition to the cosine basis hypothesis, we compare ourselves with first hitting time model DeepFHT [6], SODEN [16], and the classical Cox regression [5], as well as its neural network version DeepSurv [11].

Our contributions are:

- We show that the integral of the cosine basis has a closed form solution and does not require a solver to compute the cumulative hazard (see Section 2.1);
- We show PD progression is age biased, but predictive models' performance increases substantially when considering clinical history (see Table 1);
- We compare several non proportional and patient specific methods and conclude predicting PD progression benefits from these assumptions (see Section 4).

2 Methods

Let $\mathbf{z} \in \mathbb{R}^K$ be the per-patient context embedding produced by the two-branch encoder (see Section 3.1), and let $\{\omega_k\}_{k=1}^K$ be a fixed grid of angular frequencies, shared across patients and not

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
2026 Slovene Conference on Artificial Intelligence, Ljubljana, Slovenia
© 2026 Copyright held by the owner/author(s).
https://doi.org/10.70314/is.2026.skui.52

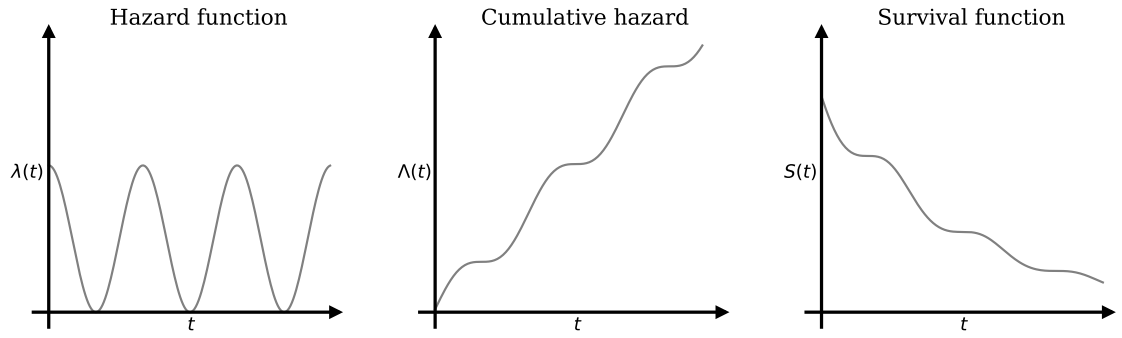

Figure 1: PD is a progressive disease that often has multiple peaks, almost like an oscillatory disease. This motivates the use of the cosine basis for the hazard function.

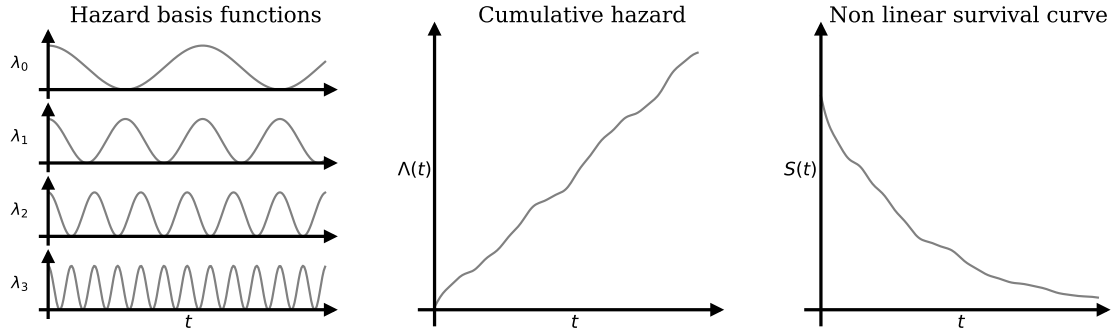

Figure 2: Definition of the survival curve with a basis composed of multiple cosine functions.

learned. The hazard is

$$\lambda(t \mid \mathbf{z}) = \sum_{k=1}^{K} a_k(\mathbf{z})\big(1 + \cos(\omega_k t)\big), \tag{2}$$

$$a_k(\mathbf{z}) = \text{softplus}\big(\mathbf{z}\big), \tag{3}$$

such that only the amplitudes depend on the patient. Because $1 + \cos(\cdot) \in [0, 2]$ and $a_k(\mathbf{z})$ is strictly positive, then $\lambda(t \mid \mathbf{z}) \geq 0$ everywhere, with no admissibility constraint to check; and because each term integrates analytically, so does the sum as

$$\Lambda(t \mid \mathbf{z}) = \int_0^t \lambda(s \mid \mathbf{z})\, \mathrm{d}s = \sum_{k=1}^{K} a_k(\mathbf{z}) \left(t + \frac{\sin(\omega_k t)}{\omega_k}\right), \tag{4}$$

with $\Lambda(0 \mid \mathbf{z}) = 0$ and $\mathrm{d}\Lambda/\mathrm{d}t = \lambda$ exactly, giving $S(t \mid \mathbf{z}) = \exp[-\Lambda(t \mid \mathbf{z})]$ at any t without a solver. An illustration of this multiple cosine formulation is shown in Figure 2

2.1 Closed Form Solution in Log Time Basis

Sampling the basis in raw normalised time weights all parts of the window equally, which is a poor match to durations that are strongly right-skewed and heavily censored. We therefore evaluate the cosines at a standardised log-time argument,

$$u(t) = \frac{\log(1 + t) - \mu}{\sigma}, \tag{5}$$

where μ and σ are the mean and standard deviation of $\log(1 + t)$ over each fold's *training* durations only.

The integral remains exact. Substituting $\tau = e^{\sigma u + \mu} - 1$, so $\mathrm{d}\tau = \sigma \mathrm{d}u + \tau\sigma \mathrm{d}u$, and using $\int e^{\sigma u} \cos(\omega u)\, \mathrm{d}u = e^{\sigma u}\big(\sigma \cos(\omega u) + \omega \sin(\omega u)\big)/(\sigma^2 + \omega^2)$ gives the antiderivative

$$F_k(t) = \frac{(\sigma + t\sigma)\,\big(\sigma \cos(\omega_k u(t)) + \omega_k \sin(\omega_k u(t))\big)}{\sigma^2 + \omega_k^2}, \tag{6}$$

from which

$$\lambda(t \mid \mathbf{z}) = \sum_k a_k\big(1 + \cos(\omega_k u(t))\big), \tag{7}$$

$$\Lambda(t \mid \mathbf{z}) = \sum_k a_k\big(t + F_k(t) - F_k(0)\big), \tag{8}$$

with $u(0) = -\mu/\sigma$. Differentiating Eq. (6) confirms $\mathrm{d}F_k/\mathrm{d}t = \cos(\omega_k u(t))$ exactly, so $\lambda = \mathrm{d}\Lambda/\mathrm{d}t$ still holds and $\Lambda(0) = 0$ by the subtraction of $F_k(0)$. No special case is needed at $\omega_k = 0$: the denominator is strictly positive and the DC component reduces to $F_k(t) - F_k(0) = t$, recovering the same $2a_k t$ contribution as before.

3 Experimental Setting

Disease severity in the Parkinson's Precision Medicine Initiative cohort (PPMI) [12, 13] is quantified using the HY scale, a standard ordinal instrument for staging Parkinson's disease. In this work the modified HY score recorded in the Part 3 table (NHY) is used, and the state space is represented as the six integer stages $\{0, 1, 2, 3, 4, 5\}$. We choose to model transitions $2 \longrightarrow 3$ and $3 \longrightarrow 4$, because these are where the models perform above random. The

summary for 2 → 3 is 835 patients, 169 transitioned to 3, and 666 censored patients. For transition 3 → 4, we have 106 patients, where 64 transitioned to 4 HY, and 42 are censored. In total there are 79.8% and 39.6% patients censored in transitions 2 → 3 and 3 → 4, respectively. The median time is 692 weeks in 2 → 3 and 213 weeks in 3 → 4. The distribution of transition is shown in the Kaplan-Meier curve in Figure 3.

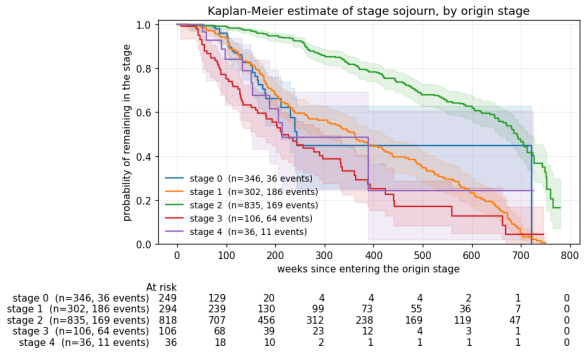

Figure 3: Kaplan–Meier stage-sojourn estimates by origin stage, with pointwise 95% confidence bands and numbers at risk.

Frequency grid. Time is normalised by a period P equal to the longest observed follow-up (780 weeks for stage 2, 746 for stage 3), so the model works in $t \approx [0, 1]$ and the cosine arguments stay small enough to avoid precision loss. The $K = 64$ frequencies are placed uniformly on $[0, 2\pi/P]$, i.e. $\omega_k = (2\pi/P)\,\kappa_k$ with κ_k evenly spaced on $[0, 1]$. We set $K = 64$ to match the 64-unit width used throughout the network and because 64 is a standard resolution for discretising the time axis in neural survival models [3, 1]; since every frequency completes at most one cycle over follow-up, a larger K would only refine the spectrum within the same smooth family rather than add flexibility to the hazard. This is not a harmonic series: the *highest* frequency completes exactly one cycle over the observation window and every other completes less. In linear time the family is therefore restricted to smooth, slowly varying hazards with at most one turning point over follow-up. Given a median sojourn of 692 weeks out of stage 2, a hazard oscillating several times within the window would be difficult to justify clinically; the restriction is deliberate. Under the log-time reparameterisation, the same frequencies are spaced uniformly in $u(t)$, which compresses late follow-up and so admits more than one turning point in t.

Amplitude initialisation. With $K = 64$ amplitudes summed, initialisation matters. At zero biases each softplus$(0) \approx 0.69$ and the summed hazard would start near $0.69K \approx 44$, driving $S(t) = \exp(-\Lambda)$ to numerical zero for every patient on the first forward pass and stalling training. Each bias is therefore initialised to $\log(e^{1/K} - 1)$, so $a_k \approx 1/K$ and $\sum_k a_k \approx 1$, giving a baseline $\Lambda(t) \approx 2t$ in normalised units.

3.1 Model Architecture

We first encode the longitudinal MDS-UPDRS history with an LSTM and the demographics with an MLP. Their embeddings are then concatenated and passed to a layer that computes the parameters for the distribution being used. The inputs are the $T \times 65 \times 7$ clinical tensor and the 22-dimensional demographic vector constructed. The outputs of both branches are concatenated and fed to a final layer that gives a prediction for the parameters of the distribution being modelled by the output.

Clinical branch. At each visit, the seven-state encoding of every Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) item is embedded into 64 dimensions. A shared per-feature LSTM processes the 65 item sequences independently, producing 65 hidden states of size 64 at each timestep. These states are flattened and projected to a single 64-dimensional visit representation. The representation at the final valid sequence position is selected using the recorded sequence length and therefore summarises the available pre-origin clinical history.

Demographics branch. The 22 demographic inputs are passed through two 64-unit layers with ReLU activation, layer normalisation, and dropout, producing a 64-dimensional embedding. Layer normalisation is used because it does not depend on batch-level statistics and is therefore suitable for the event-enriched batches used during training [2].

Hyperparameters. For all neural-network-based survival models considered in this work, we run a 5-fold cross validation with a single common configuration: two hidden layers of 64 units each, the AdamW optimiser with a learning rate of $1e-3$ and weight decay of $1e-4$, a batch size of 128, and dropout of 0.1. The learning rate follows a reduce-on-plateau schedule with factor 0.5, patience of 3 epochs, and a minimum value of $1e-6$. Early stopping uses a patience of 20 epochs, and each batch contains at least 16 observed events.

4 Results

In transition 2 → 3, the best model in terms of C-index was DeepFHT, which means this model ranked the times of individuals the best. The second best was SODEN. However, in terms of IBS, which tells us how calibrated the model is, the best model was Cox (demog+measures), while the second best was Cox (only demog). This result is not that meaningful since Cox ranks the times randomly with C index 0.5438. In terms of INBLL, the best model was again Cox (only demog), while DeepFHT came second best. Overall, DeepFHT was the best model altogether, reaching good results in terms of the various metrics analyzed. It ranks the times well and is fairly calibrated. As for transition 3 → 4, the best model is DeepFHT with 0.6957 and, in second, Cox (demog+measures) achieves 0.6644. These results are very good, however in terms of IBS, the best model achieved 0.1718 with it being a random ranking model Cox (only demog). The second best calibrated model was DeepFHT with 0.1828. As for INBLL, the best model was Cox (only demog) with 0.5472, while DeepFHT achieved similar metric with 0.5689.

The first observation we make is that the difference between a linear Cox model and a non linear DeepSurv is not clear. In terms of C-index both perform close to random. The calibration is good, as shown by the IBS and INBLL metrics. But in both transitions, the models fail to predict correctly the target. This suggests demographics features do not contain the true signal to predict time to next stage of HY. Second, as shown in Table 1, the models with demographic features plus MDS-UPDRS measures achieve a much better performance. The difference between DeepSurv with and without measures is of +0.0739 in favour of including measures. This corresponds to a 12% increase in performance. More complex models with different assumptions to proportional hazards, like DeepFHT, achieve good results for both transitions. This suggests that MDS-UPDRS history is important for this task and the hazard shape is different for each patient.

	Transition 2 → 3			Transition 3 → 4		
	C-index	IBS	INBLL	C-index	IBS	INBLL
Cox-linear (only demog) [5]	0.5438	<u>0.1366</u>	**0.4165**	0.5222	**0.1718**	**0.5472**
DeepSurv (only demog) [11]	0.5642	0.1388	0.4235	0.4554	0.1877	0.6081
DeepSurv (demog+measures) [11]	0.6529	**0.1331**	0.5042	<u>0.6644</u>	0.1919	1.2007
DeepFHT (demog+measures) [6]	**0.6947**	0.1410	<u>0.4402</u>	**0.6957**	<u>0.1828</u>	<u>0.5689</u>
SODEN (demog+measures) [16]	<u>0.6746</u>	0.1409	0.4543	0.6328	0.2604	1.1543
Cosine (demog+measures)	0.6538	0.1478	0.4580	0.6269	0.2877	1.0530

Table 1: Models' performances on prediction of time to the next stage of the HY scale in PD.

In demographics only models, the most important feature in terms of effect magnitude was age with 0.237 log-hazard units, while other demographical features such as race (≈ 0.050), sex (≈ -0.020), and sexual orientation (≈ 0.100). The MDS-UPDRS features' effect magnitudes are small (roughly -0.009 to $+0.016$) and spread across all parts rather than concentrated in one feature/part. The largest hazard-increasing effects combine axial and lower-limb motor items, swallowing difficulty (`NP2SWAL`, +0.016), rest tremor amplitude (`NP3RTARU`, +0.015), toe tapping and leg agility on the right (`NP3TTAPR`, `NP3LGAGR`), with dopamine dysregulation (`NP1DDS`) and anxiety (`NP1ANXS`) among the largest non-motor contributors. The largest hazard-decreasing items are dominated by upper-limb motor and axial measures, pronation-supination movements of the right hand and body bradykinesia (`NP3PRSPR`, `NP3BRADY`, both -0.009), difficulty rising from a chair (`NP2RISE`), and cognitive impairment (`NP1COG`). Even though the effect size is much larger for age, giving the model clinical MDS-UPDRS data increases substantially the performance of the model, suggesting that age acts more as a bias and clinical data contains the true signal of the time to the next stage.

5 Conclusion

Across the four survival heads, the most reproducible pattern concerns the inputs rather than the mathematical form of the hazard. Sparse pre-origin history in transition 2 → 3 makes demographic fusion important for discrimination, whereas richer history in transition 3 → 4 allows the clinical branch to stand on its own and makes the demographic branch consistently detrimental to pooled ranking and probabilistic accuracy. The repeated reversal across recurrent Cox, DeepFHT, SODEN, and cosine models indicates that this is primarily a property of the stage-specific data rather than an idiosyncrasy of one head. DeepFHT achieves higher discriminative performance, but the proposed cosine hazard basis remains highly attractive because its closed-form antiderivative allows exact, instantaneous cumulative hazard computation without requiring numerical ODE solvers or iterative integration.

Acknowledgements

DC is supported by EU's HE MSCA COFUND Programme grant agreement No.101081355-SMASH and by the Republic of Slovenia. DC, VG, DG, and AS are supported by EU project 22HLT05 MAIBAI and ARIS project P2-0209. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the EU or ERA. Neither the EU nor the granting authority can be held responsible for them. Data used in this article was obtained from the PPMI database, RRID:SCR_006431.

References

[1] Abdallah Alabdallah et al. 2025. CoxSE: exploring the potential of self-explaining neural networks with cox proportional hazards model for survival analysis. *Knowledge-Based Systems.*
[2] Jimmy Lei Ba et al. 2016. Layer normalization. *arXiv.*
[3] George H. Chen. 2024. Survival kernets: scalable and interpretable deep kernel survival analysis with an accuracy guarantee. *JMLR.*
[4] J Andrés Christen and Francisco Javier Rubio. 2025. On harmonic oscillator hazard functions. *Statistics & Probability Letters.*
[5] David R Cox. 1972. Regression models and life-tables. *Journal of the royal statistical society.*
[6] Alessio Cristofoletto et al. 2025. Neural diffusion processes for physically interpretable survival prediction. *arXiv.*
[7] Fabio Luis De Mello et al. 2021. Metaparametric neural networks for survival analysis. *IEEE Transactions on Neural Networks and Learning Systems.*
[8] Vincenzo Dentamaro et al. 2024. Enhancing early Parkinson's disease detection through multimodal deep learning and explainable AI: insights from the PPMI database. *Scientific Reports.*
[9] Raphael T. Gerraty et al. 2023. Machine learning within the Parkinson's progression markers initiative: review of the current state of affairs. *Frontiers in Aging Neuroscience.*
[10] Margaret M. Hoehn et al. 1967. Parkinsonism: onset, progression, and mortality. *Neurology.*
[11] Jared L Katzman et al. 2018. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology.*
[12] Kenneth Marek et al. 2011. The Parkinson progression marker initiative (PPMI). *Progress in Neurobiology.*
[13] Kenneth Marek et al. 2018. The Parkinson's progression markers initiative (PPMI)— establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology.*
[14] Xuehan Ren et al. 2021. Prognostic modeling of parkinson's disease progression using early longitudinal patterns of change. *Movement Disorders.*
[15] Mohammad Samsul Alam et al. 2025. Dynamic prediction using functional latent trait joint models for multivariate longitudinal outcomes: an application to parkinson's disease. *Statistics in medicine.*
[16] Weijing Tang et al. 2022. SODEN: a scalable continuous-time survival model through ordinary differential equation networks. *JMLR.*
[17] Y. J. Zhao et al. 2010. Progression of Parkinson's disease as evaluated by Hoehn and Yahr stage transition times. *Movement Disorders.*