



**MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME TABANLI
SAĞKALIM MODELLERİNİN KARŞILAŞTIRILMASI:
AÇIKLANABİLİR YAPAY ZEKA YAKLAŞIMI**

Esra AKAYDIN GÜLTÜRK

Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

**Tez Danışmanı
Doç. Dr. Emek GÜLDOĞAN**

Doktora Tezi-2026

**T.C.
İNÖNÜ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME TABANLI SAĞKALIM
MODELLERİNİN KARŞILAŞTIRILMASI: AÇIKLANABİLİR YAPAY ZEKA
YAKLAŞIMI**

Esra AKAYDIN GÜLTÜRK

Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

Doktora Tezi

**Tez Danışmanı
Doç. Dr. Emek GÜLDOĞAN**

**Tez Jüri Üyeleri
Doç. Dr. Emek GÜLDOĞAN
Prof. Dr. Cemil ÇOLAK
Prof. Dr. Burcu Kayhan TETİK
Doç.Dr. Feyza İNCEOĞLU
Doç.Dr.Mehmet Onur KAYA**

**MALATYA
2026**

T.C.
İNÖNÜ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

ETİK ve YAPAY ZEKÂ KULLANIM BEYANI

02 /06 / 2026

İnönü Üniversitesi Lisansüstü Tez Yazım Kurallarına uygun olarak, “Doç. Dr. Emek GÜLDOĞAN ” danışmanlığında hazırlayıp sunduğum “Makine Öğrenmesi ve Derin Öğrenme Tabanlı Sağlık Modellerinin Karşılaştırılması: Açıklanabilir Yapay Zeka Yaklaşımı” Doktora tezimde:

1. ETİK ve BİLİMSEL SORUMLULUKLARIM:

- Tezimde yer alan verileri, bilgileri ve belgeleri akademik ve etik kurallar çerçevesinde elde ettiğimi, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Çalışmada yararlandığım tüm kaynaklara bilimsel kurallara uygun şekilde atıfta bulunduğumu ve kaynakça gösterimde APA 7. baskı kurallarına uyduğumu,
- Tezimin özgün olduğunu, çalışma ve yazım sürecinde patent ve telif haklarını ihlal edici hiçbir davranışta bulunmadığımı,
- Aksi bir durumda doğabilecek tüm hukuki ve akademik sonuçları kabul ettiğimi,

2. YAPAY ZEKÂ KULLANIMINA İLİŞKİN BEYANIM:

() Tez yazım sürecinde yapay zekâ tabanlı araçlardan **faydalanmadım**.

(X) Tez yazım sürecinde yapay zekâ tabanlı araçlardan **faydalandım**.

Bu durumda:

- Yapay zekâ araçlarını yalnızca bilimsel üretim sürecini desteklemek amacıyla kullandığımı,
 - Kullanımın hipotez geliştirme, teorik tartışma ve yorumlama gibi üst düzey akademik becerileri içermediğini,
 - Kullanımın tezin özgünlüğüne, bilimsel etik kurallarına ve ilgili mevzuata aykırı olmadığını,
 - Usulsüz kullanımlara (kaynak göstermeksizin metin üretimi, yanıltıcı kaynak gösterimi, otomatik içeriklerin özgün eser gibi sunulması vb.) kesinlikle başvurmadığımı,
 - Kullanılan yapay zekâ araçlarının ve işlevlerinin, danışmanımın bilgisi dâhilinde, ekte verilen tabloda şeffaf biçimde belirtildiğini,
 - Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Faaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehber’de belirtilen ilkelere riayet ettiğimi,
- kabul ve taahhüt ederim.

Esra AKAYDIN GÜLTÜRK

İmza

Bu tablo, tez çalışmanızda kullanılan üretken yapay zekâ araçlarının şeffaf biçimde beyan edilmesi amacıyla hazırlanmıştır. Her satırda kullanılan aracın adı, kullanım amacı, hangi bölümde kullanıldığı ve kullanım şekli belirtilmelidir. Tablo, danışman öğretim üyesinin bilgisi dâhilinde doldurulmalı ve gerektiğinde ek belgelerle desteklenmelidir.

No	Üretken Yapay Zekâ Aracı	Kullanım Amacı	Kullanıldığı Bölüm	Kullanım Şekli
1	Claude	Uslûp İyileştirmesi	Genel Bilgiler	Dil ve Uslûp iyileştirme
2	Deepl	İngilizce	Özet, Genel Bilgiler	Çeviri
3	Grammarly	Gramer Kontrolü	Tartışma	Gramer Düzeltmesi
4				
5				

Üretken Yapay Zekâ Aracı: Kullanılan aracın adı (ör. ChatGPT, Grammarly, Zotero AI, Copilot vb.).

Kullanım Amacı: Aracın hangi amaçla kullanıldığı (ör. literatür özeti, dil kontrolü, kaynak önerisi, yazım denetimi vb.).

Kullanıldığı Bölüm: Tezin hangi bölümünde kullanıldığı (ör. Giriş, Literatür Taraması, Bulgular, Sonuç vb.).

Kullanım Şekli: Kullanım biçiminin niteliği (ör. öneri alma, metin taslağı oluşturma, dil ve imla kontrolü, kaynak listesi düzenleme, doğrudan alıntı vb.).

Bu tablo, tezde üretken yapay zekâ kullanımının şeffaf, izlenebilir ve etik kurallara uygun olmasını sağlamak için hazırlanmıştır. Araçların katkısı, tez metninin özgünlüğünü ve bilimsel sorumluluğu ortadan kaldıracak düzeyde olmamalıdır.

İÇİNDEKİLER

ÖZET	viii
ABSTRACT	ix
SİMGELER VE KISALTMALAR DİZİNİ	x
ŞEKİLLER DİZİNİ.....	xii
TABLolar DİZİNİ	xv
1. GİRİŞ	1
2. GENEL BİLGİLER	4
2.1. Sağkalım Analizi.....	4
2.1.1. Sansürleme	5
2.1.2. Kaplan-Meier Tahmincisi	6
2.1.3. Cox Orantılı Hazardlar Modeli (Cox Proportional Hazards).....	7
2.2. Makine Öğrenmesi Tabanlı Sağkalım Modelleri.....	9
2.2.1. Sağkalım Ağaçları (Classification and Regression Trees).....	9
2.2.2. Rastgele Sağkalım Ormanları (Random Survival Forests)	10
2.2.3. Gradyan Güçlendirme (Gradient Boosting).....	11
2.2.4. Aşırı Rastgele Sağkalım Ağaçları (Extremely Randomized Survival Trees)	12
2.2.5. Aşırı Gradyan Güçlendirme (eXtreme Gradient Boosting)	13
2.2.6. K-En Yakın Komşu (K-Nearest Neighbors).....	13
2.2.7. Sağkalım Destek Vektör Makineleri (Survival Support Vector Machines).....	14
2.3. Derin Öğrenme Tabanlı Sağkalım Modelleri.....	14
2.3.1. Cox Tabanlı Derin Öğrenme Sağkalım Modeli (DeepSurv)	15
2.3.2. Ayırık Zamanlı Derin Öğrenme Sağkalım Modeli (DeepHit)	16
2.4. Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence).....	17
2.4.1. SHAP (SHapley Additive exPlanations).....	18
2.4.2. Zamana Bağlı Sağkalım SHAP Açıklamaları (SurvSHAP(t)).....	19
2.4.3. Sağkalım için Yerel Yorumlanabilir Model-Agnostik Açıklamalar (SurvLIME).....	19
2.5. Model Değerlendirme Metrikleri	21
2.5.1. Uyum İndeksi (Concordance Index)	21

2.5.2. Bütünleşik Brier Skoru (Integrated Brier Score)	22
2.6. Çapraz Doğrulama ve Hiperparametre Optimizasyonu	23
2.7. İlgili Çalışmalar.....	25
3. MATERYAL VE METOT	29
3.1. Veri Seti	29
3.1.1. Ön işleme ve Dönüştürme.....	33
3.1.2. Modeller ve Optimizasyon.....	34
3.1.3. Değerlendirme.....	35
3.1.4. Yazılım Ortamı ve Kütüphaneler.....	35
4. BULGULAR.....	36
5. TARTIŞMA	105
6. SONUÇ VE ÖNERİLER.....	109
KAYNAKLAR	111
EKLER.....	120
EK-1. Özgeçmiş	120
EK-2. Etik Kurul Gerekmediğine Dair Yazı.....	121

TEŞEKKÜR

Doktora tez çalışmam boyunca bilgi ve deneyimleriyle bana yol gösteren, her aşamada desteğini esirgemeyen değerli danışman hocam Doç. Dr. Emek GÜLDOĞAN'a en içten teşekkürlerimi sunarım. Tez sürecinin her aşamasında ve doktora eğitim sürecinde bilimsel katkılarıyla beni yönlendiren kıymetli hocam Prof. Dr. Cemil ÇOLAK'a ayrıca minnettarım.

Biyoistatistik ve Tıp Bilişimi Anabilim Dalı'ndaki değerli hocalarım emekli Prof. Dr. Saim YOLOĞLU'na, Prof. Dr. Harika Gözde GÖZÜKARA BAĞ'a, Doç. Dr. Ahmet Kadir ARSLAN'a bölümdeki adlarını sayamadığım tüm akademisyenlere ve doktora arkadaşlarıma eğitim sürecim boyunca sağladıkları bilimsel destek ve katkılarından dolayı teşekkürü borç bilirim.

Tez izleme komitesi üyeliği ve tez savunma sürecindeki değerli katkıları için Prof. Dr. Burcu KAYHAN TETİK'e; tez savunma jürisinde yer alarak görüş ve önerileriyle çalışmama katkı sağlayan Doç.Dr. Feyza İNCEOĞLU'na ve Doç. Dr .Mehmet Onur KAYA'ya teşekkürlerimi sunarım.

Bu süreçte beni destekleyen ve her zaman yanımda olan değerli eşim Abdulaziz GÜLTÜRK'e ve hayatımın anlamı, yaşam enerjim olan canım çocuklarıma sonsuz teşekkürlerimi sunarım.

Esra AKAYDIN GÜLTÜRK

Malatya, 2026

ÖZET

Makine Öğrenmesi ve Derin Öğrenme Tabanlı Sağkalım Modellerinin Karşılaştırılması: Açıklanabilir Yapay Zeka Yaklaşımı

Amaç: Kolon kanseri sağkalım verisinde istatistiksel, ağaç tabanlı, ensemble, kernel ve derin öğrenme yaklaşımlarını temsil eden on farklı modelin tahmin performansının karşılaştırılması ve açıklanabilirlik yöntemleriyle karar mekanizmalarının değerlendirilmesi amaçlanmıştır.

Materyal ve metod: R survival paketindeki colon veri seti (n=888, olay=430, olay oranı=%48.4) kullanıldı. Model performansı 5×3 iç içe çapraz doğrulama ile değerlendirilmiş; C-index, Integrated Brier Score ve zamana bağlı AUC metrikleri hesaplandı. Model güvenilirliği 1000 tekrarlı bootstrap ile %95 güven aralıkları üzerinden değerlendirilmiştir.

Bulgular: İç içe çapraz doğrulamada en yüksek ortalama C-index DeepSurv (0.6569), RSF (0.6568) ve DeepHit (0.6567) modellerinde birbirine çok yakın elde edilmiş; RSF en düşük IBS (0.1743) ve en yüksek TD-AUC değerlerinden biri (0.7161) ile en dengeli performansı göstermiştir. Bootstrap analizinde ExtraSurvTrees en yüksek C-index (0.7350), IBS (0.1503) ve TD-AUC (0.8098) değerlerine ulaşmış; ancak iç içe çapraz doğrulamadaki düşük sıralaması bu sonucun iyimser yanlılığına işaret etmektedir. SHAP analizi tüm modellerde pozitif lenf düğümü sayısını en güçlü prognostik faktör olarak belirlemiştir. SurvSHAP(t) öznelik öneminin zamana bağlı değişimini, SurvLIME ise hasta düzeyinde yorumlanabilir risk profilleri ortaya koymuştur.

Sonuç: Ensemble yöntemler tutarlı üstünlük gösterirken derin öğrenme modelleri küçük-orta ölçekli verilerde kararsızlık riski taşımaktadır. Açıklanabilirlik analizleri modeller arası tutarlı prognostik örüntüleri doğrulayarak klinik karar destek uygulamalarına zemin hazırlamaktadır.

Anahtar Kelimeler: Açıklanabilir yapay zeka, Derin öğrenme, Kolon kanseri, Makine öğrenmesi, Sağkalım analizi, SHAP, SurvLIME, SurvSHAP(t)

ABSTRACT

Comparison Of Machine Learning And Deep Learning-Based Survival Models: An Explainable Artificial Intelligence Approach

Aim: This study aimed to compare the predictive performance of ten models representing statistical, tree-based, ensemble, kernel, and deep learning approaches on colon cancer survival data and to evaluate their decision mechanisms using explainability methods.

Material and method: The colon dataset from the R survival package (n=888, events=430, event rate=48.4%) was used. Model performance was evaluated using 5×3 nested cross-validation with C-index, Integrated Brier Score, and time-dependent AUC. Model reliability was assessed through 1000-iteration bootstrap with 95% confidence intervals.

Results: In nested cross-validation the highest mean C-index values were very close across DeepSurv (0.6569), RSF (0.6568), and DeepHit (0.6567); RSF showed the most balanced performance with the lowest IBS (0.1743) and one of the highest TD-AUC values (0.7161). In bootstrap analysis ExtraSurvTrees achieved the highest C-index (0.7350), IBS (0.1503), and TD-AUC (0.8098), although its low nested-CV ranking indicates the optimistic bias of this estimate. SHAP analysis identified the number of positive lymph nodes as the strongest prognostic factor across all models. SurvSHAP(t) revealed time-dependent variations in feature importance, while SurvLIME provided patient-level interpretable risk profiles.

Conclusion: Ensemble methods demonstrated consistent superiority while deep learning models carried instability risk on small-to-medium datasets. Explainability analyses confirmed consistent prognostic patterns across models, laying the groundwork for clinical decision support applications.

Keywords: Colon cancer, Deep learning, Explainable artificial intelligence, Machine learning, SHAP, Survival analysis, SurvLIME, SurvSHAP(t)

SİMGELER VE KISALTMALAR DİZİNİ

AUC	: Eğri Altındaki Alan (Area Under the Curve)
BS	: Brier Skoru (Brier Score)
CART	: Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees)
C-index	: Uyum İndeksi (Concordance Index)
CoxPH	: Cox Orantılı Hazardlar Modeli (Cox Proportional Hazards)
CV	: Çapraz Doğrulama (Cross-Validation)
DeepHit	: Ayrık Zamanlı Derin Öğrenme Sağkalım Modeli
DeepSurv	: Cox Tabanlı Derin Öğrenme Sağkalım Modeli
d_i	: i. Zamanda Olay Sayısı
ExtraSurvTrees	: Aşırı Rastgele Sağkalım Ağaçları (Extremely Randomized Survival Trees)
GA	: Güven Aralığı (Confidence Interval)
GradientBoosting	: Gradyan Güçlendirme (Gradient Boosting)
$h(t)$	: Hazard Fonksiyonu
$H(t)$	: Kümülatif Hazard Fonksiyonu
$h_0(t)$	: Temel (Baseline) Hazard Fonksiyonu
IBS	: Bütünleşik Brier Skoru (Integrated Brier Score)
IPCW	: Ters Olasılık Sansürleme Ağırlıklandırması (Inverse Probability of Censoring Weighting)
IQR	: Çeyrekler Arası Açıklık (Interquartile Range)
k	: Komşu Sayısı / Kat Sayısı
KDD	: Bilgi Keşfi Süreci (Knowledge Discovery in Databases)
KM	: Kaplan-Meier
KNN	: K-En Yakın Komşu (K-Nearest Neighbors)
$L(\beta)$	: Kısmi Olabilirlik Fonksiyonu
LASSO	: En Küçük Mutlak Küçültme ve Seçim Operatörü (Least Absolute Shrinkage and Selection Operator)
LIME	: Yerel Yorumlanabilir Model-Agnostik Açıklamalar (Local Interpretable Model-agnostic Explanations)

n	: Örneklem Büyüklüğü
n_i	: i. Zamanda Risk Altındaki Birey Sayısı
p	: Değişken Sayısı
ReLU	: Düzeltilmiş Doğrusal Birim (Rectified Linear Unit)
RSF	: Rastgele Sağkalım Ormanları (Random Survival Forests)
S(t)	: Sağkalım Fonksiyonu
$\hat{S}(t)$	: Tahmini Sağkalım Fonksiyonu
SELU	: Ölçekli Üssel Doğrusal Birim (Scaled Exponential Linear Unit)
SHAP	: Shapley Toplanır Açıklamalar (SHapley Additive exPlanations)
SurvivalSVM	: Sağkalım Destek Vektör Makineleri (Survival Support Vector Machines)
SurvLIME	: Sağkalım için Yerel Yorumlanabilir Model-Agnostik Açıklamalar
SurvSHAP(t)	: Zamana Bağlı Sağkalım SHAP Açıklamaları
SVM	: Destek Vektör Makineleri (Support Vector Machines)
TD-AUC	: Zamana Bağlı Eğri Altındaki Alan (Time-Dependent AUC)
TPE	: Ağaç Yapılı Parzen Tahmincisi (Tree-structured Parzen Estimator)
XAI	: Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence)
XGBoost	: Aşırı Gradyan Güçlendirme (eXtreme Gradient Boosting)
α	: Karmaşıklık Parametresi
β	: Regresyon Katsayısı
$\phi_j(t)$	: j. Özniteliğin Zamana Bağlı SHAP Değeri

ŞEKİLLER DİZİNİ

Şekil No	Sayfa No
Şekil 3.1. Bilgi keşfi süreci.....	29
Şekil 3.2. Genel Kaplan-Meier sağkalım eğrisi.....	32
Şekil 3.3. Tedavi gruplarına göre Kaplan-Meier eğrileri	33
Şekil 4.1. C-index kutu grafiği ve fold bazında ısı haritası	37
Şekil 4.2. Bütünleşik Brier skoru karşılaştırması	38
Şekil 4.3. Orman grafiği ve bootstrap dağılımları	38
Şekil 4.4. Tüm modellere ait kümülatif/dinamik AUC grafiği.....	39
Şekil 4.5. Modeller arası SHAP öznitelik önem ısı haritası	40
Şekil 4.6. CoxPH modeline ait SHAP özet grafiği	41
Şekil 4.7. RSF modeline ait SHAP özet grafiği	42
Şekil 4.8. CART modeline ait SHAP özet grafiği	43
Şekil 4.9. GradientBoosting modeline ait SHAP özet grafiği	44
Şekil 4.10. ExtraSurvTrees modeline ait SHAP özet grafiği	45
Şekil 4.11. XGBoost modeline ait SHAP özet grafiği	46
Şekil 4.12. KNN modeline ait SHAP özet grafiği	47
Şekil 4.13. SurvivalSVM modeline ait SHAP özet grafiği	48
Şekil 4.14. DeepSurv modeline ait SHAP özet grafiği.....	49
Şekil 4.15. DeepHit modeline ait SHAP özet grafiği	50
Şekil 4.16. Modeller arası SurvSHAP(t) ısı haritası.....	51
Şekil 4.17. CoxPH modeline ait global SurvSHAP(t) grafiği	52
Şekil 4.18. RSF modeline ait global SurvSHAP(t) grafiği.....	53
Şekil 4.19. CART modeline ait global SurvSHAP(t) grafiği	54
Şekil 4.20. GradientBoosting modeline ait global SurvSHAP(t) grafiği	55
Şekil 4.21. ExtraSurvTrees modeline ait global SurvSHAP(t) grafiği.....	56
Şekil 4.22. XGBoost modeline ait global SurvSHAP(t) grafiği.....	57
Şekil 4.23. KNN modeline ait global SurvSHAP(t) grafiği	58
Şekil 4.24. SurvivalSVM modeline ait global SurvSHAP(t) grafiği.....	59
Şekil 4.25. DeepSurv modeline ait global SurvSHAP(t) grafiği	60
Şekil 4.26. DeepHit modeline ait global SurvSHAP(t) grafiği	61
Şekil 4.27. CoxPH modeline ait SurvSHAP(t) agregasyon grafiği.....	62

Şekil 4.28. RSF modeline ait SurvSHAP(t) agregasyon grafiği.....	63
Şekil 4.29. CART modeline ait SurvSHAP(t) agregasyon grafiği	64
Şekil 4.30. GradientBoosting modeline ait SurvSHAP(t) agregasyon grafiği	65
Şekil 4.31. ExtraSurvTrees modeline ait SurvSHAP(t) agregasyon grafiği.....	66
Şekil 4.32. XGBoost modeline ait SurvSHAP(t) agregasyon grafiği.....	67
Şekil 4.33. KNN modeline ait SurvSHAP(t) agregasyon grafiği	68
Şekil 4.34. SurvivalSVM modeline ait SurvSHAP(t) agregasyon grafiği	69
Şekil 4.35. DeepSurv modeline ait SurvSHAP(t) agregasyon grafiği.....	70
Şekil 4.36. DeepHit modeline ait SurvSHAP(t) agregasyon grafiği	70
Şekil 4.37. CoxPH modeline ait SurvSHAP(t) yüksek riskli hasta profili	71
Şekil 4.38. RSF modeline ait SurvSHAP(t) yüksek riskli hasta profili.....	72
Şekil 4.39. CART modeline ait SurvSHAP(t) yüksek riskli hasta profili	73
Şekil 4.40. GradientBoosting modeline ait SurvSHAP(t) yüksek riskli hasta profili	74
Şekil 4.41. ExtraSurvTrees modeline ait SurvSHAP(t) yüksek riskli hasta profili.....	75
Şekil 4.42. XGBoost modeline ait SurvSHAP(t) yüksek riskli hasta profili.....	76
Şekil 4.43. KNN modeline ait SurvSHAP(t) yüksek riskli hasta profili	77
Şekil 4.44. SurvivalSVM modeline ait SurvSHAP(t) yüksek riskli hasta profili.....	78
Şekil 4.45. DeepSurv modeline ait SurvSHAP(t) yüksek riskli hasta profili.....	79
Şekil 4.46. DeepHit modeline ait SurvSHAP(t) yüksek riskli hasta profili	80
Şekil 4.47. CoxPH modeline ait SurvSHAP(t) düşük riskli hasta profili.....	81
Şekil 4.48. RSF modeline ait SurvSHAP(t) düşük riskli hasta profili	82
Şekil 4.49. CART modeline ait SurvSHAP(t) düşük riskli hasta profili.....	83
Şekil 4.50. GradientBoosting modeline ait SurvSHAP(t) düşük riskli hasta profili	84
Şekil 4.51. ExtraSurvTrees modeline ait SurvSHAP(t) düşük riskli hasta profili	85
Şekil 4.52. XGBoost modeline ait SurvSHAP(t) düşük riskli hasta profili	86
Şekil 4.53. KNN modeline ait SurvSHAP(t) düşük riskli hasta profili.....	87
Şekil 4.54. SurvivalSVM modeline ait SurvSHAP(t) düşük riskli hasta profili	88
Şekil 4.55. DeepSurv modeline ait SurvSHAP(t) düşük riskli hasta profili.....	89
Şekil 4.56. DeepHit modeline ait SurvSHAP(t) düşük riskli hasta profili	90
Şekil 4.57. CoxPH modeline ait SurvLIME en yüksek riskli hasta profili	94
Şekil 4.58. CoxPH modeline ait SurvLIME yüksek riskli hasta profili	95
Şekil 4.59. CoxPH modeline ait SurvLIME orta riskli hasta profili	96
Şekil 4.60. CoxPH modeline ait SurvLIME düşük riskli hasta profili	97
Şekil 4.61. CoxPH modeline ait SurvLIME en düşük riskli hasta profili	98

Şekil 4.62. CoxPH modeline ait SurvLIME katsayıları ısı haritası (5 gözlem)	99
Şekil 4.63. RSF modeline ait SurvLIME en yüksek riskli hasta profili	100
Şekil 4.64. RSF modeline ait SurvLIME orta riskli hasta profili	101
Şekil 4.65. RSF modeline ait SurvLIME düşük riskli hasta profili.....	102
Şekil 4.66. RSF modeline ait SurvLIME en düşük riskli hasta profili	103
Şekil 4.67. RSF modeline ait SurvLIME katsayıları ısı haritası (4 gözlem)	104



TABLÖLAR DİZİNİ

Tablo No	Sayfa No
Tablo 3.1. KDD aşamalarının bu çalışmadaki karşılıkları	30
Tablo 3.2. Veri setindeki değişkenlerin tanımları	31
Tablo 3.3. Tedavi grupları arası ikili log-rank karşılaştırmaları	33
Tablo 3.4. Ölçeklendirme stratejisi	34
Tablo 3.5. Modeller.....	34
Tablo 3.6. Optimizasyon yöntemi	34
Tablo 4.1. İç içe çapraz doğrulama (5×3) sonuçları.....	36
Tablo 4.2. Bootstrap %95 güven aralıkları (B=1000).....	36
Tablo 4.3. Normalize SHAP değerleri	39
Tablo 4.4. SurvSHAP(t) agregasyonu yapılmış değerler	50
Tablo 4.5. Kullanılan modeller ve parametreleri	91
Tablo 4.6. SurvLIME katsayıları: CoxPH modeli (5 gözlem)	92
Tablo 4.7. SurvLIME katsayıları: RSF modeli (4 gözlem).....	92

1. GİRİŞ

Kanser, dünya genelinde en önemli halk sağlığı sorunlarından biri olmaya devam etmektedir. Dünya Sağlık Örgütü verilerine göre 2022 yılında yaklaşık 20 milyon yeni kanser vakası ve 9.7 milyon kansere bağlı ölüm gerçekleşmiştir (1). Kolorektal kanser, küresel ölçekte en sık görülen üçüncü kanser türü ve kansere bağlı ölümlerde ikinci sırada yer almaktadır (2). Türkiye'de ise kolorektal kanser erkeklerde üçüncü, kadınlarda ikinci en sık görülen kanser türüdür (3). Erken tanı ve tedavi yöntemlerindeki gelişmelere rağmen kolorektal kanserin beş yıllık sağkalım oranı evre ve tedavi yaklaşımına göre %14 ile %91 arasında değişmektedir (4). Bu geniş aralık, prognostik faktörlerin doğru değerlendirilmesinin ve bireyselleştirilmiş tedavi stratejilerinin geliştirilmesinin kritik önemini vurgulamaktadır.

Sağkalım analizi, bir olayın gerçekleşme süresini modelleyen istatistiksel yöntemlerin genel adıdır. Tıbbi araştırmalarda bu olay çoğunlukla ölüm, hastalığın tekrarlama veya belirli bir komplikasyonun gelişmesi olarak tanımlanmaktadır. Kaplan-Meier tahmincisi ve Cox orantılı hazardlar modeli, sağkalım analizinin temel yapı taşları olarak onlarca yıldır klinik araştırmalarda yaygın şekilde kullanılmaktadır (5, 6). Ancak bu klasik yöntemler, değişkenler arasındaki doğrusal olmayan ilişkileri ve karmaşık etkileşimleri yakalamada sınırlı kalabilmektedir (7).

Son yıllarda makine öğrenmesi ve derin öğrenme yöntemleri, sağkalım analizi alanında giderek artan bir ilgiyle araştırılmaktadır. Rastgele sağkalım ormanları, gradyan güçlendirme sağkalım modelleri, destek vektör makineleri ve derin öğrenme tabanlı yaklaşımlar olan DeepSurv (Cox tabanlı derin öğrenme sağkalım modeli) ve DeepHit (Ayrık Zamanlı Derin Öğrenme Sağkalım Modeli) gibi yöntemler, klasik istatistiksel modellerle karşılaştırılarak değerlendirilmektedir (8-12). Bu yöntemler özellikle yüksek boyutlu veri setlerinde ve doğrusal olmayan ilişkilerin mevcut olduğu durumlarda geleneksel yöntemlere göre üstün performans sergileyebilmektedir (13, 14). Bununla birlikte, makine öğrenmesi ve derin öğrenme modellerinin kara kutu yapısı, klinik karar verme süreçlerinde güvenilirlik ve şeffaflık açısından ciddi endişeler doğurmaktadır (15).

Açıklanabilir yapay zeka (Explainable Artificial Intelligence, XAI), kara kutu modellerin tahminlerinin arkasındaki nedenleri anlaşılır hale getirmeyi amaçlayan yöntemler bütünüdür. Shapley toplanır açıklamalar (SHapley Additive exPlanations, SHAP) (16), yerel yorumlanabilir model-agnostik açıklamalar (Local Interpretable

Model-agnostic Explanations, LIME) (17) ve bu yöntemlerin sağkalım analizine uyarlanmış versiyonları olan SurvSHAP(t) (zamana bağlı sağkalım SHAP açıklamaları) (18) ve SurvLIME (sağkalım için yerel yorumlanabilir model-agnostik açıklamalar) (19), sağkalım modellerinin açıklanabilirliğini artırmak amacıyla geliştirilmiştir. SurvSHAP(t) zamana bağlı öznitelik önem değerleri üreterek her bir zaman noktasında hangi değişkenlerin sağkalım tahminini ne yönde etkilediğini göstermektedir. SurvLIME ise bireysel hasta düzeyinde lokal açıklamalar sunarak klinisyenlerin tedavi kararlarını desteklemektedir. Bu yöntemlerin sağkalım analizinde sistematik ve karşılaştırmalı olarak uygulanması, modellerin klinik pratikte güvenle kullanılabilirliğini artırma potansiyeli taşımaktadır (20).

Klinik karar destek sistemlerinde sağkalım modellerinin kullanımı, son on yılda önemli bir artış gösterdi. Amerikan Klinik Onkoloji Derneği ve Avrupa Tıbbi Onkoloji Derneği gibi uluslararası kuruluşlar, bireyselleştirilmiş tedavi kararlarında tahmin modellerinin potansiyel rolünü vurgulamıştır (117). Bununla birlikte, bu modellerin klinik pratikte yaygın olarak benimsenmesinin önündeki en büyük engellerden biri, kara kutu modellerinin karar verme süreçlerindeki şeffaflık eksikliğidir (15, 21). Klinisyenler, bir modelin neden belirli bir hastayı yüksek riskli olarak sınıflandırdığını anlamadan tedavi kararlarını bu modele dayandırmaktan kaçınmaktadırlar. Bu durum, XAI yöntemlerinin sağkalım analizine entegrasyonunun klinik açıdan neden bu kadar önemli olduğunu açıkça düşündürmektedir.

Sağkalım analizi modellerinin performans karşılaştırmasında metodolojik titizlik büyük önem taşımaktadır. Farklı çalışmalarda farklı değerlendirme stratejileri (tek bölme, k-katlı çapraz doğrulama, iç içe çapraz doğrulama), farklı performans metrikleri ve farklı veri ön işleme yaklaşımları kullanılması, sonuçların doğrudan karşılaştırılmasını zorlaştırmaktadır. İç içe çapraz doğrulama, hiperparametre optimizasyonundan kaynaklanan iyimserlik yanlılığını önlemek için altın standart olarak kabul edilmektedir (22, 23). Bootstrap güven aralıkları ise nokta tahminlerinin güvenilirliğini değerlendirmekte önemli bir araç sunmaktadır. Bu tez çalışmasında hem iç içe çapraz doğrulama hem de bootstrap analizi birlikte uygulanarak metodolojik sağlamlık sağlandı.

Literatürde makine öğrenmesi tabanlı sağkalım analizleri yaygın olmakla birlikte, çok sayıda modelin aynı veri seti üzerinde kapsamlı bir şekilde karşılaştırılması ve bu modellerin açıklanabilirlik boyutunun sistematik olarak değerlendirildiği çalışmalar sınırlıdır. Özellikle istatistiksel, ağaç tabanlı, ensemble, uzaklık/kernel tabanlı ve derin öğrenme yaklaşımlarının birlikte değerlendirildiği ve üç farklı açıklanabilirlik yöntemiyle

(SHAP, SurvSHAP(t), SurvLIME) desteklendiği bütüncül çalışmaların eksikliği dikkat çekmektedir (24, 25).

Bu tezin amacı, kolon kanseri sağkalım verisinde on farklı makine öğrenmesi ve derin öğrenme modelini (Cox orantılı hazardlar modeli (Cox Proportional Hazards, CoxPH), rastgele sağkalım ormanları (Random Survival Forests, RSF), sınıflandırma ve regresyon ağaçları (Classification and Regression Trees, CART), gradyan güçlendirme (Gradient Boosting, GradientBoosting), aşırı rastgele sağkalım ağaçları (Extremely Randomized Survival Trees, ExtraSurvTrees), aşırı gradyan güçlendirme (eXtreme Gradient Boosting, XGBoost), K-en yakın komşu (K-Nearest Neighbors, KNN), sağkalım destek vektör makineleri (Survival Support Vector Machines, SurvivalSVM), DeepSurv, DeepHit) kapsamlı performans metrikleriyle (C-index, bütünleşik Brier skoru (Integrated Brier Score, IBS), zamana bağlı eğri altındaki alan (Time-Dependent AUC, TD-AUC)) karşılaştırmak ve üç farklı açıklanabilirlik yöntemi (SHAP, SurvSHAP(t), SurvLIME) ile modellerin karar mekanizmalarını şeffaf hale getirmektir. Çalışmanın hipotezi, derin öğrenme ve ensemble yöntemlerin klasik Cox regresyonuna kıyasla anlamlı düzeyde yüksek tahmin performansı göstereceği ve açıklanabilirlik yöntemlerinin modeller arasında tutarlı öznelik önem örüntüleri ortaya koyacağı şeklindedir. Bu doğrultuda, derin öğrenme modellerinin sağkalım analizindeki potansiyelini değerlendirmek, en uygun tahmin modelini belirlemek ve açıklanabilirlik yöntemleriyle desteklenen bir tahmin çerçevesi geliştirmek hedeflendi.

2. GENEL BİLGİLER

2.1. Sağkalım Analizi

Sağkalım analizi, belirli bir olayın gerçekleşme süresini inceleyen istatistiksel yöntemler topluluğudur. Tıbbi araştırmalarda bu olay genellikle ölüm, hastalık nüksü, organ reddi veya belirli bir komplikasyonun gelişmesi olarak tanımlanır (26). Sağkalım analizi, zamana kadar olan süreyi (time-to-event) modellemekte ve sansürlü gözlemleri dikkate alabilmesi sayesinde diğer regresyon yöntemlerinden ayrılmaktadır (27). Mühendislik alanında güvenilirlik analizi, ekonomide süre modelleri ve sosyal bilimlerde olay geçmişi analizi gibi farklı adlarla da kullanılmaktadır.

Sağkalım analizinin temel bileşenleri sağkalım fonksiyonu $S(t)$, hazard fonksiyonu $h(t)$ ve kümülatif hazard fonksiyonu $H(t)$ olarak üç ana kavramla ifade edilmektedir. Sağkalım fonksiyonu $S(t)$, bir bireyin t zamanından daha uzun süre hayatta kalma olasılığını ifade etmektedir (Denklem 2.1). Hazard fonksiyonu $h(t)$, belirli bir zamanda anlık risk oranını temsil etmektedir (Denklem 2.2). Kümülatif hazard fonksiyonu $H(t)$ ise 0'dan t 'ye kadar birikmiş toplam riski göstermektedir (Denklem 2.3). Sağkalım fonksiyonu ile kümülatif hazard arasındaki ilişki Denklem (2.4)'te verilmiştir (28). Bu üç fonksiyon birbirine matematiksel olarak dönüştürülebilir ve sağkalım analizinin temelini oluşturmaktadır.

$$S(t) = P(T > t) \quad (2.1)$$

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad (2.2)$$

$$H(t) = -\ln S(t) \quad (2.3)$$

$$S(t) = \exp -H(t) \quad (2.4)$$

Sağkalım verileri parametrik, yarı-parametrik ve parametrik olmayan yöntemlerle modellenebilmektedir. Parametrik yöntemler (üstel, Weibull, log-normal, log-lojistik dağılımlar) sağkalım süresinin belirli bir olasılık dağılımına uyduğunu varsaymaktadır (29). Bu varsayım doğru olduğunda parametrik modeller en verimli tahminleri üretmektedir; ancak varsayımın ihlali ciddi yanlışlığa yol açabilmektedir. Yarı-parametrik Cox modeli, temel hazard fonksiyonunu belirtilmemiş bırakarak bu sorunu kısmen çözmektedir. Parametrik olmayan yöntemler (Kaplan-Meier, Nelson-Aalen) ise herhangi bir dağılım varsayımı gerektirmemektedir (27).

Sağkalım analizinin tarihsel gelişimi incelendiğinde, ilk çalışmaların 19. yüzyılda aktüerya biliminde mortalite tablolarının hesaplanmasıyla başladığı görülmektedir. Modern sağkalım analizi ise 1950'li ve 1970'li yıllarda Kaplan-Meier tahmincisi ve Cox orantılı hazardlar modelinin önerilmesiyle şekillenmiştir (5, 6). 2000'li yıllardan itibaren makine öğrenmesi ve derin öğrenme yöntemlerinin sağkalım analizine entegrasyonu, alanı yeni bir boyuta taşımıştır. Günümüzde sağkalım analizi tıp, mühendislik, finans, sosyal bilimler ve pazarlama gibi pek çok alanda yaygın olarak kullanılmaktadır.

Sağkalım verilerinin modellenmesinde karşılaşılan temel zorluklar arasında sansürlü gözlemler, zamana bağlı kovaryatlar, yarışan riskler, tekrarlayan olaylar ve bilgilendirici sansürleme yer almaktadır. Yarışan riskler modellemesi, birden fazla olay türünün mümkün olduğu durumlarda (örneğin bir hastanın kanser nedeniyle veya kardiyovasküler hastalık nedeniyle ölebileceği durumlar) önem kazanmaktadır. Fine ve Gray (109) yarışan riskler için alt dağılım hazard modelini önermişlerdir. Tekrarlayan olaylar modellemesi ise Andersen-Gill, Wei-Lin-Weissfeld ve Prentice-Williams-Peterson gibi genişletilmiş Cox modelleri ile ele alınmaktadır (120).

2.1.1. Sansürleme

Sansürleme, sağkalım analizinin en ayırt edici özelliğidir ve bir gözlem için ilgilenilen olayın çalışma süresi içinde gözlenemediği durumu ifade etmektedir. Sansürleme üç türde sınıflandırılmaktadır: sağdan sansürleme, soldan sansürleme ve aralık sansürleme (29). Klinik çalışmalarda en yaygın tür sağdan sansürlemedir ve hastanın çalışma süresi sona erdiğinde hâlâ hayatta olması, takipten çıkması veya ilgisiz bir nedenle kaybedilmesi durumlarında ortaya çıkmaktadır (32).

Sansürleme mekanizmasının bağımsız (non-informative) olduğu varsayımı, çoğu sağkalım analizi yönteminin temelini oluşturmaktadır. Bu varsayım, sansürleme zamanının sağkalım süresiyle ilişkisiz olduğunu belirtmektedir. Bu varsayımın ihlali, sağkalım tahminlerinde yanlılığa yol açabilmektedir (33). Bilgilendirici sansürleme durumlarında ortak fragilite modelleri veya yarışan riskler modelleri gibi alternatif yaklaşımlar kullanılmaktadır (34). Sansürleme oranı veri setinin bilgi içeriğini doğrudan etkileyen önemli bir faktördür; yüksek sansürleme oranları model performansını ve tahmin güvenilirliğini azaltabilmektedir.

Sansürleme oranı, veri setindeki bilgi miktarını ve model performansını doğrudan etkileyen kritik bir faktördür. Yüksek sansürleme oranları modelin öğrenebileceği olay bilgisini azaltmakta ve performans tahminlerinin güvenilirliğini düşürmektedir. Genel bir

kural olarak sansürleme oranının yüzde 50'nin altında tutulması önerilmektedir; ancak klinik çalışmalarda bu eşiğin aşılması sıklıkla görülmektedir. Sansürleme oranının yüksek olduğu durumlarda ters olasılık sansürleme ağırlıklandırması (Inverse Probability of Censoring Weighting, IPCW) tabanlı performans metrikleri ve sansürlemeye duyarlı kayıp fonksiyonları tercih edilmelidir (27, 32).

Sağdan sansürleme kendi içinde Tip I (sabit zamanlı), Tip II (sabit sayılı) ve rastgele sansürleme olarak alt türlere ayrılmaktadır. Tip I sansürlemede çalışma önceden belirlenmiş bir zamanda sona ermekte ve o ana kadar olay yaşamamış tüm bireyler sansürlenmektedir. Tip II sansürlemede çalışma önceden belirlenmiş sayıda olay gözlenene kadar devam etmektedir. Klinik çalışmalarda en yaygın tür olan rastgele sansürlemede ise bireyler farklı zamanlarda takipten çıkabilmektedir (35).

2.1.2. Kaplan-Meier Tahmincisi

Kaplan-Meier (KM) tahmincisi, sağkalım fonksiyonunu parametrik olmayan bir şekilde tahmin eden en yaygın kullanılan yöntemdir. 1958 yılında Kaplan ve Meier tarafından önerilmiştir (5). KM tahmincisi, her bir olay zamanında koşullu sağkalım olasılıklarının çarpımı olarak hesaplanmaktadır (Denklem 2.5). Bu formülde n_i ilgili zamanda risk altındaki birey sayısını, d_i ise olay sayısını ifade etmektedir (36).

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i} \quad (2.5)$$

KM tahmincisinin güçlü yönleri arasında parametrik varsayım gerektirmemesi, sansürlü gözlemleri doğal olarak ele alması ve görsel olarak yorumlanabilir sağkalım eğrileri üretmesi sayılabilir. Gruplar arası sağkalım farklılıklarının test edilmesinde log-rank testi, Gehan-Breslow-Wilcoxon testi ve Tarone-Ware testi gibi istatistiksel testler kullanılmaktadır. Log-rank testi eşit ağırlıklı, Gehan-Breslow-Wilcoxon testi erken zamanlara daha fazla ağırlık vermektedir (37, 38). KM tahmincisinin temel sınırlılığı, kovaryatların etkisini doğrudan modelleyememesidir; bu nedenle çok değişkenli analizlerde Cox regresyon modeli tercih edilmektedir (39).

KM tahmincisinin varyansı Greenwood formülü ile tahmin edilmektedir. Bu formül, güven aralıklarının hesaplanmasında kullanılmaktadır. KM eğrisi için nokta bazlı güven aralıkları ve güven bantları farklı yaklaşımlarla elde edilebilmektedir. Log-transformasyon tabanlı güven aralıkları sağkalım olasılığının $[0, 1]$ aralığında kalmasını garanti ettiği için tercih edilmektedir (40). Risk tablosu (number at risk table) KM

eğrisinin altında gösterilmekte ve belirli zaman noktalarında kaç bireyin hâlâ risk altında olduğu bilgisini sunmaktadır.

KM tahmincisinin klinik araştırmalardaki kullanımı birkaç önemli noktaya dikkat edilmesini gerektirmektedir. Birincisi, KM eğrisinin yorumlanmasında medyan sağkalım süresi ve belirli zaman noktalarındaki sağkalım oranları en yaygın kullanılan özet istatistiklerdir. Medyan sağkalım süresi, sağkalım fonksiyonunun yüzde 50'ye düştüğü zaman noktası olarak tanımlanmaktadır; ancak yüksek sağkalım oranlarına sahip veri setlerinde medyan süre hesaplanamayabilmektedir (26, 36). İkincisi, KM eğrisinin kuyruk bölümünde (geç zamanlarda) risk altındaki birey sayısının azalması nedeniyle tahminlerin güvenilirliği düşmektedir; bu nedenle risk tablosunun KM eğrisi ile birlikte sunulması standart uygulamadır. Üçüncüsü, çok sayıda grubun karşılaştırılmasında çoklu karşılaştırma düzeltmeleri (Bonferroni, Holm gibi) uygulanması gerekebilmektedir (121, 122).

2.1.3. Cox Orantılı Hazardlar Modeli (Cox Proportional Hazards)

Cox orantılı hazardlar modeli, 1972 yılında Sir David Cox tarafından önerilmiş olup sağkalım analizinde en yaygın kullanılan çok değişkenli regresyon modelidir. Modelin hazard fonksiyonu Denklem (2.6)'da verildi. Bu formülde $h_0(t)$ temel hazard fonksiyonunu, β katsayıları değişkenlerin etkisini ve X değişken değerlerini göstermektedir (41).

$$h(t|X) = h_0(t) \cdot \exp \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (2.6)$$

Cox modeli yarı-parametrik bir yapıya sahiptir; temel hazard fonksiyonu $h_0(t)$ için herhangi bir parametrik dağılım varsayımı yapılmamakta, regresyon katsayıları ise kısmi olabilirlik (partial likelihood) yöntemiyle tahmin edilmektedir (Denklem 2.7) (42). Modelin temel varsayımı olan orantılı hazardlar varsayımı, hazard oranlarının zamandan bağımsız olduğunu belirtmektedir. Bu varsayımın geçerliliği Schoenfeld artıkları testi, log(-log) grafikleri ve zaman bağımlı kovaryatlar ile değerlendirilmektedir (43). Orantılı hazardlar varsayımının ihlali durumunda stratifiye Cox modeli, zamana bağlı katsayılar veya parametrik hızlandırılmış yaşam süresi modelleri alternatif olarak kullanılabilir (35).

$$L(\beta) = \prod_{i=1}^D \frac{\exp \beta' X_i}{\sum_{j \in R_i} \exp \beta' X_j} \quad (2.7)$$

Cox modelinin avantajları arasında yorumlanabilirliğinin yüksek olması, hazard oranlarının doğrudan klinik anlam taşıması ve sağlam bir istatistiksel temele sahip olması

sayılabilir. Bununla birlikte, modelin doğrusal varsayımı, yüksek boyutlu verilerde çoklu doğrusal bağımlılık riski ve kovaryatlar arasındaki karmaşık etkileşimleri yakalayamaması gibi sınırlılıkları mevcuttur (44). Bu sınırlılıklar, makine öğrenmesi tabanlı sağkalım modellerinin geliştirilmesinin temel motivasyonunu oluşturmaktadır.

Cox modelinin genelleştirilmiş versiyonları arasında stratifiye Cox modeli, zamana bağlı kovaryatlı Cox modeli ve penalize Cox modeli (Ridge, LASSO, Elastic Net düzenlileştirmeli) yer almaktadır. Penalize Cox modeli, yüksek boyutlu verilerde değişken seçimi yapabilmekte ve aşırı öğrenmeyi kontrol edebilmektedir. En küçük mutlak küçültme ve seçim operatörü (Least Absolute Shrinkage and Selection Operator, LASSO) düzenlileştirmesi bazı katsayıları sıfıra çekerek otomatik değişken seçimi gerçekleştirmektedir (45). Elastic Net her iki yaklaşımın bir kombinasyonunu sunmaktadır (46). Cox modelinde temel hazard fonksiyonu $h_0(t)$ belirtilmemiş olmasına rağmen, sağkalım fonksiyonunun tahmini için Breslow tahmincisi kullanılabilir. Breslow tahmincisi (Denklem 2.8), sağkalım fonksiyonu üretmeyen makine öğrenmesi modelleri için de uygulanabilir olması nedeniyle önem taşımaktadır (47).

$$\hat{H}_0(t) = \sum_{t_i \leq t} \frac{d_i}{\sum_{j \in R_i} \exp \beta' X_j} \quad (2.8)$$

Sağkalım analizinde model karşılaştırması yapılırken dikkate alınması gereken bir diğer önemli nokta, modellerin farklı çıktı türleri üretmesidir (24, 25). Bazı modeller doğrudan sağkalım fonksiyonu $S(t)$ tahmini üretirken, bazıları yalnızca risk skoru veya hazard oranı üretmektedir. Cox modeli, RSF, CART, GradientBoosting ve ExtraSurvTrees doğrudan kümülatif hazard fonksiyonu ve sağkalım fonksiyonu tahmini üretebilmektedir (6, 8). Buna karşılık XGBoost, KNN, SurvivalSVM, DeepSurv ve DeepHit gibi modeller yalnızca risk skorları üretmekte olup sağkalım fonksiyonunun elde edilmesi için ek adımlar gerekmektedir (65, 68).

Makine öğrenmesi modellerinin sağkalım analizindeki performansı, veri setinin özelliklerine bağlı olarak önemli ölçüde değişebilmektedir. Örneklem boyutu, değişken sayısı, değişken türleri, sansürleme oranı, olay oranı ve değişkenler arası korelasyon yapısı gibi veri özelliklerinin her biri model performansını etkileyen faktörlerdir. Küçük örneklem boyutlarında basit modeller (Cox, CART) genellikle daha kararlı sonuçlar üretirken, büyük örneklem boyutlarında karmaşık modeller (RSF, XGBoost, derin öğrenme) doğrusal olmayan ilişkileri yakalayarak üstün performans sergileyebilmektedir (48, 49). Değişkenler arası etkileşimlerin güçlü olduğu veri setlerinde ağaç tabanlı ensemble modellerin avantajlı olduğu gözlenmektedir (8, 55).

2.2. Makine Öğrenmesi Tabanlı Sağkalım Modelleri

Makine öğrenmesi, veriden örüntü çıkaran ve öğrenme yoluyla tahmin yapabilen algoritmalar bütünüdür (50). Sağkalım analizinde makine öğrenmesi yöntemlerinin uygulanması, klasik istatistiksel modellerin sınırlılıklarını aşmak amacıyla 1990'lı yıllardan itibaren araştırılmaktadır. Sağkalım analizine uyarlanmış makine öğrenmesi modelleri, sansürlü gözlemleri ele alabilecek şekilde kayıp fonksiyonlarının ve bölme kriterlerinin özel olarak tasarlanmasını gerektirmektedir (51). Sağkalım analizinde yaygın kayıp fonksiyonları arasında Cox kısmi olabilirliği, log-rank bölme kriteri, sıralama kayıp fonksiyonu ve kalibrasyon kayıp fonksiyonu yer almaktadır.

Makine öğrenmesi tabanlı sağkalım modellerinin temel avantajları şunlardır: doğrusal olmayan ilişkileri yakalayabilme, değişkenler arası etkileşimleri otomatik olarak modelleyebilme, yüksek boyutlu veri setlerinde etkili olabilme ve parametrik dağılım varsayımı gerektirmeme (52). Bununla birlikte, bu modellerin kara kutu yapısı, aşırı öğrenme riski, büyük örneklem gereksinimleri ve hiperparametre ayarlamasının karmaşıklığı önemli dezavantajlar olarak değerlendirilmektedir (53).

Ensemble öğrenme, birden fazla temel öğrenicinin tahminlerinin birleştirilmesiyle tek bir öğreniciden daha güçlü bir tahmin modeli oluşturmayı amaçlayan bir yaklaşımdır (54). Ensemble yöntemler iki ana stratejiye dayanmaktadır: bagging (Bootstrap Aggregating) ve boosting. Bagging yaklaşımında birbirinden bağımsız eğitilen modellerin tahminleri ortalamalanarak varyans azaltılmaktadır; RSF ve ExtraTrees bu kategoriye girmektedir. Boosting yaklaşımında modeller sıralı olarak eğitilmekte ve her yeni model önceki modelin hatalarına odaklanmaktadır; gradyan güçlendirme ve XGBoost bu kategorinin temsilcileridir (55, 66). Yanlılık-varyans ayrıştırması perspektifinden bagging ağırlıklı olarak varyansı, boosting ise yanlılığı azaltmaktadır.

2.2.1. Sağkalım Ağaçları (Classification and Regression Trees)

Sınıflandırma ve regresyon ağaçları (Classification and Regression Trees, CART) Breiman ve arkadaşları tarafından 1984 yılında önerilmiştir (57). Sağkalım ağaçları, olay zamanı verilerini özyinelemeli olarak bölen (recursive partitioning) yapıdadır. Bölme kriteri olarak log-rank istatistiği veya Harrell'in C-index'i kullanılmaktadır (37, 58). Her yaprak düğüm, o alt grupta tahmin edilen Kaplan-Meier sağkalım fonksiyonunu temsil etmektedir.

LeBlanc ve Crowley sağkalım ağaçlarını sansürlü verilere uyarlayarak ilk sistematik çerçeveyi oluşturmuşlardır (59). Sağkalım ağacı oluşturma süreci üç aşamadan oluşmaktadır: büyütme (growing), budama (pruning) ve son ağaç seçimi. Büyütme aşamasında ağaç mümkün olan en derin seviyeye kadar büyütülmektedir. Budama aşamasında aşırı öğrenmeyi önlemek amacıyla maliyet-karmaşıklık budaması (cost-complexity pruning) uygulanmaktadır. Son ağaç seçiminde ise çapraz doğrulama ile optimum karmaşıklık parametresi belirlenmektedir (60).

Ağaç tabanlı modellerin en önemli avantajı yüksek yorumlanabilirliktir; karar kuralları doğrudan görselleştirilebilir ve klinik açıdan anlaşılır ifadeler sağlanmaktadır. Bununla birlikte, tek bir karar ağacı yüksek varyansa sahip olma eğilimindedir ve küçük veri değişikliklerine karşı hassastır. Bu sınırlılık, ensemble yöntemlerin geliştirilmesine motivasyon sağlamıştır (54).

Sağkalım ağaçlarının bir diğer önemli uygulaması koşullu çıkarım ağaçlarıdır (conditional inference trees). Hothorn ve arkadaşları (61) tarafından önerilen bu yaklaşım, bölme değişkeninin seçiminde istatistiksel hipotez testleri kullanarak seçim yanlılığını azaltmaktadır. Geleneksel CART algoritmasında çok düzeyli kategorik değişkenler ve sürekli değişkenler bölme kriteri açısından avantajlı olabilmekteyken, koşullu çıkarım ağaçları bu yanlılığı düzeltmektedir.

Ağaç tabanlı sağkalım modellerinde budama stratejileri model karmaşıklığını kontrol etmede kritik rol oynamaktadır. Maliyet-karmaşıklık budaması (cost-complexity pruning), bir karmaşıklık parametresi α ile ağaç boyutu ve tahmin hatası arasındaki dengeyi optimize etmektedir (57, 59). Çapraz doğrulama ile optimal α değeri belirlenmekte ve bu değere karşılık gelen alt ağaç seçilmektedir. Minimum düğüm boyutu (nodesize) ise yaprak düğümlerdeki minimum gözlem sayısını belirleyerek aşırı öğrenmeyi kontrol etmektedir.

2.2.2. Rastgele Sağkalım Ormanları (Random Survival Forests)

Rastgele sağkalım ormanları (Random Survival Forests, RSF), rastgele ormanlar yaklaşımının sağkalım analizine uyarlanmış versiyonudur (62). İlk olarak Ishwaran ve arkadaşları tarafından 2008 yılında önerilmiştir (8). RSF, bootstrap örnekleme ile çok sayıda sağkalım ağacı oluşturmakta ve her bölme noktasında rastgele seçilen bir değişken alt kümesini değerlendirmektedir. RSF algoritmasının adımları şu şekilde özetlenebilir: (i) orijinal veriden bootstrap ile B adet örneklem çekilir, (ii) her örneklem üzerinde bir sağkalım ağacı oluşturulurken her düğümde rastgele seçilen p adet değişken arasından en

iyi bölme belirlenir, (iii) ağaçlar minimum düğüm boyutuna ulaşana kadar budanmadan büyütülür, (iv) B ağacın sağkalım fonksiyonu tahminlerinin ortalaması alınarak ensemble tahmini elde edilir (8).

RSF'nin bölme kriteri olarak log-rank istatistiği, log-rank skoru, kümülatif hazard fonksiyonu veya C-index kullanılabilmektedir (63). Her ağacın tahmin ettiği sağkalım fonksiyonlarının ortalaması alınarak nihai ensemble tahmini elde edilmektedir. RSF'nin temel avantajları arasında doğrusal olmayan ilişkileri yakalayabilmesi, orantılı hazardlar varsayımı gerektirmemesi, değişken etkileşimlerini otomatik olarak modelleyebilmesi ve out of bag hata tahmini sayesinde ayrı bir doğrulama setine ihtiyaç duymaması sayılabilir (51).

RSF'nin performansı ağaç sayısı (ntree), her bölmede değerlendirilen değişken sayısı (mtry) ve minimum düğüm boyutu (nodesize) gibi hiperparametrelere bağlıdır. Genellikle ağaç sayısının artırılması performansı iyileştirmekte ancak belirli bir noktadan sonra hesaplama maliyetini gereksiz yere artırmaktadır. Ampirik çalışmalar RSF'nin özellikle orta ve büyük örneklem boyutlarında Cox modeline kıyasla üstün performans sergilediğini göstermiştir (56, 64). RSF'nin önemli bir çıktısı olan değişken önem (Variable Importance, VIMP) değerleri, permütasyon tabanlı yaklaşımla hesaplanmaktadır: bir değişkenin değerleri rastgele karıştırıldığında model performansındaki düşüş, o değişkenin önemini yansıtmaktadır (63).

2.2.3. Gradyan Güçlendirme (Gradient Boosting)

Gradyan güçlendirme (Gradient Boosting, GradientBoosting), zayıf öğrencilerin sıralı olarak birleştirilmesiyle güçlü bir tahmin modeli oluşturan bir ensemble öğrenme yaklaşımıdır. Friedman tarafından 2001 yılında önerilen genel gradyan güçlendirme çerçevesi (55), sağkalım analizine Ridgeway tarafından uyarlanmıştır (9). Sağkalım gradyan güçlendirme modeli, Cox kısmi olabilirliğinin negatif gradyanını kayıp fonksiyonu olarak kullanmaktadır (61).

Her iterasyonda bir önceki modelin artık hatalarına odaklanan yeni bir ağaç eklenmekte ve öğrenme oranı (learning rate) ile bu ağacın katkısı kontrol edilmektedir. Düzenleştirme mekanizmaları olarak öğrenme oranı küçültme (shrinkage), alt örnekleme (subsampling) ve ağaç derinliği sınırlama kullanılmaktadır (65). Bu mekanizmalar aşırı öğrenmeyi önlemekte ve modelin genelleme kapasitesini artırmaktadır. Öğrenme oranı ile ağaç sayısı arasında ters bir ilişki bulunmaktadır; düşük

öğrenme oranı genellikle daha fazla ağaç gerektirmekte ancak daha iyi genelleme performansı sağlamaktadır.

Gradyan güçlendirme ailesi içinde scikit-survival kütüphanesindeki GradientBoostingSurvivalAnalysis uygulaması ile XGBoost kütüphanesinin sağkalım modu arasında önemli implementasyon farklılıkları bulunmaktadır. GradientBoostingSurvivalAnalysis doğrudan sağkalım fonksiyonu tahmini üretebilirken, XGBoost yalnızca risk skorları vermektedir. Ayrıca XGBoost ikinci dereceden Taylor yaklaşımı ile daha agresif optimizasyon yapmakta ve L1/L2 düzenleme terimleri doğrudan amaç fonksiyonuna entegre edilmektedir (64, 66). Bu farklılıklar, iki modelin aynı gradyan güçlendirme ailesine ait olmasına rağmen farklı performans profilleri sergilemesine yol açabilmektedir.

Ensemble modellerde ağaç çeşitliliği (tree diversity) tahmin performansı için kritik bir faktördür. Bagging yaklaşımında çeşitlilik bootstrap örnekleme ve rastgele öznelik seçimi ile sağlanmaktadır (62). Boosting yaklaşımında ise çeşitlilik, ardışık modellerin birbirini tamamlamasından kaynaklanmaktadır (55). ExtraTrees modeli, hem örnekleme hem de bölme noktası seçiminde rastgelelik ekleyerek en yüksek çeşitliliğe ulaşmaktadır. Yüksek çeşitlilik genel olarak daha düşük varyansa ve daha iyi genelleme performansına yol açmaktadır (54); ancak aşırı rastgelelik yanlılığı artırabilmektedir. Bu denge, hiperparametre ayarlaması ile optimize edilmektedir.

2.2.4. Aşırı Rastgele Sağkalım Ağaçları (Extremely Randomized Survival Trees)

Aşırı rastgele sağkalım ağaçları (Extremely Randomized Survival Trees, ExtraSurvTrees), Geurts ve arkadaşları tarafından 2006 yılında önerilen aşırı rastgele ağaçlar (Extremely Randomized Trees, ExtraTrees) yönteminin sağkalım analizine uyarlanmış versiyonudur (118). Bu yöntem rastgele ormanlardan iki temel noktada ayrılır: birincisi, bootstrap örnekleme yerine tüm eğitim verisi kullanılmakta; ikincisi, bölme noktaları belirlenirken en iyi bölme noktası aranmak yerine rastgele bölme noktaları seçilmektedir. Bu ek rastgelelik, modelin varyansını daha da azaltmaktadır.

ExtraTrees yaklaşımının sağkalım analizine uyarlanması, bölme kriteri olarak sağkalıma özgü metriklerin (log-rank istatistiği gibi) kullanılmasıyla gerçekleştirilmektedir. Bu yöntemin avantajları arasında rastgele ormanlardan daha düşük varyans üretmesi, hesaplama maliyetinin daha az olması ve aşırı öğrenmeye karşı daha dirençli olması yer almaktadır (118). Dezavantaj olarak, yanlılığın rastgele

ormanlara kıyasla biraz daha yüksek olabilmesi gösterilmektedir. Pölsterl (64) scikit-survival kütüphanesinde ExtraSurvivalTrees uygulamasını sunarak bu yöntemin Python ekosisteminde kullanılabilirliğini sağlamıştır.

2.2.5. Aşırı Gradyan Güçlendirme (eXtreme Gradient Boosting)

Aşırı gradyan güçlendirme (eXtreme Gradient Boosting, XGBoost), Chen ve Guestrin (66) tarafından 2016 yılında önerilmiştir ve gradyan güçlendirme algoritmasının ölçeklenebilir ve yüksek performanslı bir uygulamasıdır. XGBoost, ikinci dereceden Taylor açılımı kullanarak kayıp fonksiyonunu yaklaşık olarak optimize etmekte ve düzenleme terimleri (L1 ve L2) ekleyerek modelin karmaşıklığını kontrol etmektedir (66).

Sağkalım analizinde XGBoost, CoxPH kayıp fonksiyonu (survival:cox) ile kullanılabilir. Bu kayıp fonksiyonu, Cox kısmi olasılığının negatifini minimize etmekte ve risk skorları üretmektedir. XGBoost'un avantajları arasında yüksek hesaplama verimliliği, eksik değer yönetimi, paralel işleme desteği ve güçlü düzenleme mekanizmaları bulunmaktadır (67). Sağkalım analizinde XGBoost'un doğrudan sağkalım fonksiyonu üretmemesi bir sınırlılık oluşturmakta ve bu durumda Breslow tahmincisi ile temel hazard fonksiyonunun elde edilmesi gerekmektedir (47).

2.2.6. K-En Yakın Komşu (K-Nearest Neighbors)

K-en yakın komşu (K-Nearest Neighbors, KNN) algoritması, tahmin yapılacak gözlemin öznitelik uzayında en yakın k komşusunun bilgisine dayanarak tahmin üreten parametrik olmayan bir yöntemdir (68). Sağkalım analizinde KNN yaklaşımı, tahmin yapılacak bireyin en yakın komşularının sağkalım verilerini kullanarak Kaplan-Meier sağkalım fonksiyonunu yerel olarak tahmin etmektedir (69).

KNN sağkalım modelinde uzaklık metriği (Öklid, Manhattan, Minkowski), komşu sayısı (k) ve ağırlıklandırma şeması (düzgün veya uzaklığa ters orantılı) temel hiperparametrelerdir. Küçük k değerleri yüksek varyansa, büyük k değerleri yüksek yanlılığa yol açmaktadır. Özniteliklerin farklı ölçeklerde olması durumunda uzaklık hesaplaması etkileneceğinden ölçeklendirme (standardizasyon veya normalizasyon) zorunludur (70). Boyutluluğun laneti (curse of dimensionality) yüksek boyutlu veri setlerinde KNN'nin performansını önemli ölçüde düşürebilmektedir.

KNN sağkalım tahmini için önerilen çeşitli yaklaşımlar mevcuttur. Beran (110) tarafından önerilen kernel tabanlı yaklaşımda, komşuların katkıları uzaklıklarına göre

ağırlıklandırılmaktadır. Lowsky ve arkadaşları (69) ise KNN'nin sağkalım olasılığı tahmininde kullanımını sistematize ederek farklı mesafe metriklerinin performansa etkisini incelemişlerdir. Öklid uzaklığı en yaygın kullanılan metrik olmakla birlikte, kategorik değişkenler içeren veri setlerinde Gower uzaklığı daha uygun olabilmektedir.

2.2.7. Sağkalım Destek Vektör Makineleri (Survival Support Vector Machines)

Destek vektör makineleri (Support Vector Machines, SVM), istatistiksel öğrenme kuramına dayanan ve yapısal risk minimizasyonu ilkesini kullanan güçlü bir sınıflandırma ve regresyon yöntemidir (71). Sağkalım analizine uyarlanmış SVM modelleri, sansürlü veriler üzerinde çalışabilecek şekilde özel kayıp fonksiyonları ile formüle edildi. Van Belle ve arkadaşları (10) sağkalım destek vektör makinelerinin kapsamlı bir çerçevesini sunmuşlardır.

Sağkalım SVM'de amaç, gözlemler arasındaki sıralama ilişkisini koruyacak şekilde bir hiper düzlem bulmaktır. Kernel fonksiyonları (doğrusal, radyal tabanlı fonksiyon, polinom) aracılığıyla doğrusal olmayan ilişkiler modellenebilmektedir. Sağkalım SVM'nin avantajları arasında güçlü genelleme kapasitesi, kernel esnekliği ve yapısal risk minimizasyonu ile aşırı öğrenme kontrolü sayılabilir. Sınırlılıkları arasında büyük veri setlerinde ölçeklenebilirlik sorunları, hiperparametre seçiminin karmaşıklığı ve doğrudan sağkalım fonksiyonu üretememesi yer almaktadır (72).

Sağkalım SVM'de farklı formülasyonlar mevcuttur. Sıralama yaklaşımı (ranking approach) gözlem çiftleri arasındaki sıralama ilişkisini korumayı amaçlamaktadır. Regresyon yaklaşımı ise sağkalım süresinin doğrudan tahmin edilmesini hedeflemektedir. Hibrit yaklaşımlar her iki formülasyonu birleştirmektedir. Van Belle ve arkadaşları (72) farklı formülasyonların sistematik karşılaştırmasını sunarak sıralama yaklaşımının genel olarak daha tutarlı sonuçlar ürettiğini gösterdi. C parametresi (düzenleştirme) ve kernel parametreleri (gamma, derece) sağkalım SVM'nin performansını doğrudan etkileyen kritik hiperparametrelerdir.

2.3. Derin Öğrenme Tabanlı Sağkalım Modelleri

Derin öğrenme, çok katmanlı yapay sinir ağları aracılığıyla veriden hiyerarşik temsiller öğrenen makine öğrenmesi yaklaşımıdır (73). Sağkalım analizinde derin öğrenme modelleri, özellikle karmaşık doğrusal olmayan ilişkileri yakalama ve yüksek boyutlu veri setlerinde üstün performans sergileme potansiyeli nedeniyle artan bir ilgiyle

araştırılmaktadır (74). Bu modeller, sansürlü gözlemleri ele alabilecek şekilde özelleştirilmiş kayıp fonksiyonları gerektirmektedir.

Derin öğrenme tabanlı sağkalım modellerinin tarihsel gelişimi incelendiğinde, ilk çalışmaların Cox modelinin sinir ağı genellemesi olarak ortaya çıktığı görülmektedir. Faraggi ve Simon (75) 1995 yılında tek gizli katmanlı bir sinir ağını Cox kısmi olasılık kayıp fonksiyonuyla birleştirmişlerdir. GPU tabanlı hesaplama altyapısının gelişmesi ve modern eğitim tekniklerinin (Adam optimizer, batch normalization, dropout) ortaya çıkmasıyla birlikte derin öğrenme sağkalım modelleri 2018 yılından itibaren hızla ilerlemiştir (76).

Derin öğrenme sağkalım modelleri, kayıp fonksiyonlarına göre üç ana kategoride sınıflandırılabilir. Birinci kategori, Cox kısmi olasılığına dayanan modeller (DeepSurv, Cox-nnet) olup orantılı hazardlar varsayımını sürdürmektedir. İkinci kategori, ayrık zamanlı yaklaşımlar (DeepHit, DRSA) olup zaman eksenini ayrıştırılarak olasılık kütle fonksiyonları tahmin edilmektedir. Üçüncü kategori ise parametrik dağılım tabanlı yaklaşımlar (Deep Survival Machines) olup sağkalım dağılımının parametrelerini doğrudan tahmin etmektedir (74). Bu tez çalışmasında birinci ve ikinci kategoriden birer temsilci model (DeepSurv ve DeepHit) seçildi.

2.3.1. Cox Tabanlı Derin Öğrenme Sağkalım Modeli (DeepSurv)

DeepSurv, Katzman ve arkadaşları tarafından 2018 yılında önerilmiştir ve CoxPH modelinin derin sinir ağı genellemesidir (11). Model, çok katmanlı bir ileri beslemeli sinir ağı kullanarak kovaryatlardan log-risk fonksiyonunu tahmin etmektedir. Kayıp fonksiyonu olarak Cox kısmi olasılığının negatifi kullanılmaktadır. Modelin çıktısı olan risk skoru Denklem (2.11)'de verildi. Bu formülde $f_{\theta}(x)$ sinir ağının çıktısını temsil etmektedir.

$$\hat{h}_{\theta}(x) = \exp f_{\theta}(x) \quad (2.11)$$

DeepSurv'ün mimarisi giriş katmanı, gizli katmanlar, aktivasyon fonksiyonları (genellikle ölçekli üssel doğrusal birim (Scaled Exponential Linear Unit, SELU) veya düzeltilmiş doğrusal birim (Rectified Linear Unit - ReLU)), dropout düzenlemesi ve çıkış katmanından oluşmaktadır. Model, batch normalizasyon ve öğrenme oranı planlaması gibi modern derin öğrenme tekniklerinden yararlanmaktadır (76). DeepSurv'ün avantajları arasında kovaryatlar arasındaki karmaşık doğrusal olmayan etkileşimleri yakalayabilmesi ve tedavi öneri sistemi olarak kullanılabilmesi sayılabilir.

Sınırlılıkları arasında orantılı hazardlar varsayımını sürdürmesi, büyük örneklem gerektirmesi ve kara kutu yapısı yer almaktadır (77).

2.3.2. Ayrık Zamanlı Derin Öğrenme Sağkalım Modeli (DeepHit)

DeepHit, Lee ve arkadaşları tarafından 2018 yılında önerilmiştir ve sağkalım analizinde parametrik varsayımlardan tamamen bağımsız bir derin öğrenme yaklaşımıdır (12). Model, olay zamanı dağılımını ayrıklaştırarak doğrudan olasılık kütle fonksiyonunu (probability mass function) tahmin etmektedir. Bu yaklaşım, zaman ekseninin K aralığa bölünmesiyle her aralıkta olay olasılığının tahmin edilmesine dayanmaktadır.

DeepHit'in kayıp fonksiyonu iki bileşenden oluşmaktadır: log-olabilirlik bileşeni ve sıralama bileşeni (ranking loss). Log-olabilirlik bileşeni, gözlenen olay zamanlarının olasılıklarını maksimize etmektedir. Sıralama bileşeni ise daha yüksek riskli bireylerin daha erken olay zamanlarına sahip olmasını sağlayacak şekilde tahminleri düzenlemektedir (78). DeepHit'in en önemli avantajı orantılı hazardlar varsayımı gerektirmemesi ve yarışan riskleri doğal olarak modelleyebilmesidir. Bununla birlikte, zaman ayrıklaştırması çözünürlüğünün belirlenmesi ve küçük veri setlerinde aşırı öğrenme riski önemli sınırlılıklar olarak değerlendirilmektedir (79).

Derin öğrenme tabanlı sağkalım modellerinin eğitiminde karşılaşılan pratik zorluklar arasında hiperparametre uzayının genişliği, küçük veri setlerinde aşırı öğrenme riski, hesaplama maliyeti ve tekrarlanabilirlik sorunları yer almaktadır. Katman sayısı, nöron sayısı, aktivasyon fonksiyonu, dropout oranı, öğrenme oranı, batch boyutu ve epoch sayısı gibi çok sayıda hiperparametre eş zamanlı olarak ayarlanmalıdır. Küçük ve orta ölçekli klinik veri setlerinde derin öğrenme modellerinin geleneksel yöntemlere göre anlamlı bir performans artışı sağlayıp sağlayamadığı tartışmalı bir konudur (76).

Derin öğrenme sağkalım modellerinin karşılaştırmasında dikkat edilmesi gereken önemli metodolojik noktalar bulunmaktadır. Birincisi, farklı kayıp fonksiyonlarına sahip modellerin doğrudan karşılaştırılabilmesi için ortak metriklerin (C-index, IBS, TD-AUC) kullanılması gerekmektedir. İkincisi, rastgele tohum değerlerinin sonuçları etkilemesi nedeniyle birden fazla rastgele başlangıç ile deneylerin tekrarlanması önerilmektedir. Üçüncüsü, sağkalım fonksiyonu üretmeyen modellerde Breslow tahmincisi veya benzeri yöntemlerle sağkalım fonksiyonunun elde edilmesi gerekmektedir (47).

Sağkalım analizinde derin öğrenme modellerinin eğitiminde kullanılan optimizasyon algoritmaları da performansı etkileyen önemli bir faktördür. Adam (Adaptive Moment Estimation) optimizasyonu, öğrenme oranını her parametre için ayrı

ayrı uyarlayan ve momentum mekanizması kullanan bir yöntem olup derin öğrenme sağkalım modellerinde en yaygın kullanılan optimizasyon algoritmasıdır (73). Öğrenme oranı planlaması (learning rate scheduling) stratejileri arasında adım azaltma (step decay), kosinüs tavlama (cosine annealing) ve doğrusal ısınma (linear warmup) yer almaktadır. Early stopping mekanizması ise doğrulama setindeki performansın iyileşmeyi durdurduğu noktada eğitimi sonlandırarak aşırı öğrenmeyi önlemektedir.

Derin öğrenme modellerinde ağ mimarisinin belirlenmesi (katman sayısı, nöron sayısı, aktivasyon fonksiyonu) hiperparametre optimizasyonunun en kritik bileşenidir. Sığ ağlar (bir veya iki gizli katman) küçük veri setlerinde yeterli olabilirken, derin ağlar karmaşık ilişkileri modellemede avantaj sağlayabilmektedir. Bununla birlikte, aşırı derin ağlar küçük veri setlerinde aşırı öğrenme riskini artırmaktadır (73, 80). SELU (Scaled Exponential Linear Unit) aktivasyon fonksiyonu, self-normalizing özelliği sayesinde batch normalization gerektirmeden kararlı eğitim sağlayabilmektedir ve DeepSurv modelinde yaygın olarak tercih edilmektedir.

2.4. Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence)

Açıklanabilir yapay zeka (Explainable Artificial Intelligence, XAI), yapay zeka modellerinin karar verme süreçlerini insan tarafından anlaşılır hale getirmeyi amaçlayan yöntemler ve araştırma alanıdır (21). Sağlık bilimlerinde klinik karar destek sistemlerinin güvenilirliği, hesap verebilirliği ve düzenleyici uyumu açısından açıklanabilirlik kritik önem taşımaktadır (81). Avrupa Birliği Genel Veri Koruma Yönetmeliği kapsamında bireylerin otomatik karar verme süreçleri hakkında açıklama alma hakkı, XAI'nin önemini daha da artırmıştır (82).

XAI yöntemleri kapsamlarına göre global ve lokal açıklamalar olarak ikiye ayrılmaktadır. Global açıklamalar modelin genel davranışını, lokal açıklamalar ise bireysel tahminlerin nedenlerini ortaya koymaktadır (83). Model bağımlılığına göre ise model-spesifik ve model-agnostik yöntemler olarak sınıflandırılmaktadır. Model-agnostik yöntemler herhangi bir modele uygulanabilirken, model-spesifik yöntemler belirli model mimarilerine özgüdür (84). Sağkalım analizinde XAI'nin uygulanması, zamana bağlı yapı nedeniyle ek zorluklar içermektedir ve bu alana özgü yöntemler geliştirilmiştir (85).

Açıklanabilirlik kavramı, yorumlanabilirlik (interpretability) kavramından farklılık göstermektedir. Yorumlanabilirlik, bir modelin doğası gereği anlaşılır olması anlamına gelmektedir; doğrusal regresyon ve karar ağaçları doğal olarak yorumlanabilir

modellerdir. Açıklanabilirlik ise herhangi bir modelin tahminlerinin post-hoc yöntemlerle anlaşılır hale getirilmesini ifade etmektedir (83). Rudin yüksek riskli karar verme (15) süreçlerinde kara kutu modellerin açıklanması yerine doğrudan yorumlanabilir modellerin kullanılmasını savunmuştur. Bununla birlikte, karmaşık modellerin bazı durumlarda anlamlı performans artışı sağlaması ve açıklanabilirlik yöntemlerinin gelişmesi, her iki yaklaşımın birlikte değerlendirilmesinin daha pragmatik bir strateji olduğunu düşündürmektedir.

Açıklanabilirlik yöntemlerinin güvenilirliği ve kararlılığı (stability) da önemli bir araştırma konusudur. Aynı gözlem için farklı çalıştırmalarda veya farklı pertürbasyon stratejileriyle elde edilen açıklamaların tutarlı olması beklenmektedir. SHAP değerlerinin aksiyomatik temeli bu tutarlılığı teorik olarak garanti etmekte; ancak yaklaşık hesaplama yöntemlerinde (KernelSHAP gibi) örnekleme varyansı nedeniyle tutarsızlıklar gözlemlenmektedir. TreeSHAP ise ağaç tabanlı modeller için kesin değerleri hesapladığından bu sorundan etkilenmemektedir (86). SurvLIME'da ise pertürbasyon stratejisi ve kernel genişliği açıklamaların kararlılığını doğrudan etkilemektedir.

Klinisyenlerin yapay zeka destekli karar destek sistemlerine güven duyabilmesi için modelin neden belirli bir tahminde bulunduğunu anlamaları gerekmektedir (87). Açıklanabilirlik ayrıca modelin olası yanlışlıklarının tespit edilmesinde, veri kalitesinin değerlendirilmesinde ve modelin klinik bilgiyle tutarlılığının doğrulanmasında kritik rol oynamaktadır (88). Bu nedenle tez çalışmasında üç farklı açıklanabilirlik yönteminin birlikte kullanılması tercih edildi.

2.4.1. SHAP (SHapley Additive exPlanations)

SHAP, Lundberg ve Lee tarafından 2017 yılında önerilmiş olup kooperatif oyun teorisindeki Shapley değerlerine dayanan model-agnostik bir açıklanabilirlik yöntemidir (16). SHAP, her bir öz niteliğin model tahminine marjinal katkısını hesaplamaktadır. Bir öz niteliğin SHAP değeri, o öz niteliğin tüm olası öz nitelik alt kümeleri üzerindeki ortalama marjinal katkısı olarak tanımlanmaktadır (89). Shapley değerleri kooperatif oyun teorisinde adil paylaşım ilkesine dayanmakta olup, n öz nitelik için 2^n alt küme üzerinden hesaplama gerektirmektedir; bu nedenle yaklaşık algoritmalar kullanılmaktadır.

SHAP değerleri üç önemli özelliğe sahiptir: yerel doğruluk (local accuracy), eksiklik (missingness) ve tutarlılık (consistency). Yerel doğruluk özelliği, tüm öz niteliklerin SHAP değerlerinin toplamının model tahmininden temel değer çıkarılmış

haline eşit olmasını garanti etmektedir (90). SHAP'ın farklı uygulamaları bulunmaktadır: TreeSHAP ağaç tabanlı modeller için $O(TLD^2)$ karmaşıklıkta verimli hesaplama sağlamakta, KernelSHAP herhangi bir model için, DeepSHAP derin öğrenme modelleri için optimize edilmiştir (86). Sağkalım analizinde SHAP değerleri, modelin risk skoru tahminlerinin açıklanmasında kullanılmaktadır. Bununla birlikte, standart SHAP zamana bağlı bilgiyi doğrudan modelleyememektedir; bu sınırlılık SurvSHAP(t) yönteminin geliştirilmesine neden olmuştur.

2.4.2. Zamana Bağlı Sağkalım SHAP Açıklamaları (SurvSHAP(t))

SurvSHAP(t), Krzyżiński ve arkadaşları tarafından 2023 yılında önerilmiştir ve SHAP değerlerinin sağkalım analizine zamana bağlı bir genellemesidir (18). Standart SHAP değerleri yalnızca tek bir skaler tahmin için açıklama üretirken, SurvSHAP(t) her bir zaman noktasında özniteliklerin sağkalım fonksiyonu tahminine katkısını hesaplamaktadır (91).

SurvSHAP(t) değeri (Denklem 2.12), her zaman noktası t için j . özneliğin sağkalım fonksiyonu üzerindeki marjinal etkisini ifade etmektedir. Bu yaklaşım sayesinde bir özneliğin erken dönemde koruyucu, geç dönemde risk artırıcı olabileceği gibi zamana bağlı dinamik ilişkiler ortaya konabilmektedir (18). Global düzeyde SurvSHAP(t) değerleri bireysel açıklamalar üzerinden agregasyon yapılarak elde edilmektedir. Agregasyon yöntemleri arasında mutlak ortalama, standart sapma ve zamana bağlı ağırlıklı ortalama yer almaktadır.

$$\varphi_j(t) = E \left(S(t|X) - S(t|X_{-j}) \right) \quad (2.12)$$

SurvSHAP(t) yönteminin avantajları arasında zamana bağlı açıklamalar üretmesi, model-agnostik olması, SHAP'ın aksiyomatik özelliklerini koruması ve hem global hem de lokal açıklamalar sunabilmesi sayılabilir. survex R paketi ve survshap Python paketi bu yöntemin pratik uygulaması için geliştirilmiştir (91). Sınırlılıkları arasında hesaplama maliyetinin yüksek olması ve yeterli gözlem sayısı gerektirmesi yer almaktadır (92).

2.4.3. Sağkalım için Yerel Yorumlanabilir Model-Agnostik Açıklamalar (SurvLIME)

SurvLIME, Kovalev ve arkadaşları tarafından 2020 yılında önerilmiştir ve LIME yönteminin sağkalım analizine uyarlanmış versiyonudur (17, 19). SurvLIME, karmaşık bir sağkalım modelinin bireysel tahminlerini, yerel olarak yaklaşıklanmış bir CoxPH modeli ile açıklamaktadır. Açıklanacak gözlemin çevresinde pertürbe edilmiş örnekler

oluşturulmakta, bu örnekler için karmaşık modelin sağkalım fonksiyonu tahminleri alınmakta ve bu tahminlere en iyi uyan bir Cox modeli yerel olarak fit edilmektedir (91).

Elde edilen Cox katsayıları, açıklanacak gözlem için özniteliklerin yerel önemini temsil etmektedir. Pozitif katsayılar risk artırıcı, negatif katsayılar koruyucu etki anlamına gelmektedir. SurvLIME yaklaşımında pertürbasyon stratejisi ve kernel genişliği sonuçları doğrudan etkileyen kritik hiperparametrelerdir. Kernel genişliğinin küçük tutulması daha yerel ancak potansiyel olarak dengesiz açıklamalar üretirken, geniş tutulması yerelliği azaltmaktadır (92).

SurvLIME'in avantajları arasında bireysel hasta düzeyinde klinik olarak yorumlanabilir açıklamalar üretmesi, model-agnostik olması ve sonuçların hazard oranları cinsinden ifade edilebilmesi sayılabilir. Sınırlılıkları arasında yerel yaklaşıklamanın doğruluğunun komşuluk yapısına bağlı olması ve yalnızca lokal açıklamalar sunması yer almaktadır (92). SurvLIME, klinik pratikte bireysel tedavi kararlarının desteklenmesinde değerli bir araç olarak görülmektedir.

SHAP değerlerinin görselleştirilmesinde birçok standart grafik türü kullanılmaktadır. Summary plot (özet grafiği) her bir öznitelik için SHAP değerlerinin dağılımını göstermekte ve öznitelik değerlerinin renk kodlamasıyla değişkenin yönünü ortaya koymaktadır. Bar plot (çubuk grafik) mutlak ortalama SHAP değerlerini sıralayarak global öznitelik önemini sunmaktadır. Waterfall plot (şelale grafik) bireysel bir tahmin için her özneliğin katkısını basamaklı olarak göstermektedir (16, 90). Force plot ise benzer bilgiyi yatay bir gösterimle sunmaktadır.

SurvSHAP(t) değerlerinin görselleştirilmesinde zamana bağlı profiller (temporal profiles) kullanılmaktadır. Her bir öznitelik için zaman ekseninde SurvSHAP(t) değerlerinin değişimi çizgi grafiği olarak gösterilmekte ve hangi zaman aralıklarında hangi özniteliklerin baskın olduğu görsel olarak ortaya konmaktadır. Agregasyon grafikleri ise tüm gözlemler üzerinden SurvSHAP(t) değerlerinin mutlak ortalamasını veya standart sapmasını göstererek global düzeyde zamana bağlı öznitelik önemini sunmaktadır (18). Bu tez çalışmasında hem bireysel hem de global SurvSHAP(t) grafikleri tüm on model için üretilmiştir.

SurvLIME sonuçlarının yorumlanmasında Cox katsayıları cinsinden hazard oranları kullanılmaktadır. Pozitif katsayılar riski artıran faktörleri, negatif katsayılar ise koruyucu faktörleri temsil etmektedir (19). SurvLIME katsayıları yalnızca yerel olarak geçerli olduğundan, farklı risk profillerindeki hastalar için farklı katsayılar elde edilmektedir. Bu tez çalışmasında en yüksek riskli, yüksek riskli, orta riskli, düşük riskli

ve en düşük riskli hasta profillerinin SurvLIME katsayıları karşılaştırılmıştır. Farklı risk düzeylerindeki hastaların hangi faktörlere daha duyarlı olduğu bu karşılaştırma ile ortaya konmuştur.

2.5. Model Değerlendirme Metrikleri

Sağkalım modellerinin performansının güvenilir bir şekilde değerlendirilmesi, model seçimi ve karşılaştırması açısından kritik öneme sahiptir. Sansürlü verilerin varlığı, standart sınıflandırma metriklerinin doğrudan kullanılmasını engellemekte ve sağkalıma özgü performans metriklerinin geliştirilmesini gerektirmektedir (93). Model performansı iki temel boyutta değerlendirilmektedir: ayırt edicilik (discrimination) ve kalibrasyon (calibration). Ayırt edicilik, modelin yüksek ve düşük riskli bireyleri ne ölçüde doğru sıralayabildiğini; kalibrasyon ise tahmin edilen olasılıkların gerçek sonuçlarla ne ölçüde uyumlu olduğunu ölçmektedir (94).

2.5.1. Uyum İndeksi (Concordance Index)

Uyum indeksi (Concordance index, C-index), Harrell ve arkadaşları tarafından 1982 yılında önerilmiştir ve sağkalım modellerinin ayırt edici gücünü ölçen en yaygın kullanılan metriktir (58). C-index, karşılaştırılabilir gözlem çiftleri arasında modelin tahminlerinin gerçek sıralama ile ne ölçüde uyumlu olduğunu ifade etmektedir. Değeri 0.5 ile 1.0 arasında olup, 0.5 rastgele tahmini, 1.0 mükemmel ayırt ediciliği temsil etmektedir (95).

C-index hesaplamasında yalnızca karşılaştırılabilir çiftler kullanılmaktadır. Bir çift, her iki gözlemden en az birinde olay gerçekleşmişse ve daha kısa sağkalım süresine sahip gözlem daha yüksek risk skoru almışsa uyumlu sayılmaktadır (96). C-index'in sınırlılıkları arasında zamandan bağımsız bir ölçüt olması, kalibrasyonu değerlendirememesi ve sansürleme oranından etkilenebilmesi sayılabilir. Uno ve arkadaşları (96) ters olasılık sansürleme ağırlıklandırması (IPCW) tabanlı bir C-index versiyonu önererek sansürleme yanlılığını azaltmışlardır.

C-index'in hesaplanmasında farklı yaklaşımlar mevcuttur. Harrell'in orijinal C-index'i sansürlü gözlemleri karşılaştırma dışında bırakmaktadır. Gönen ve Heller (111) sansürlemeye daha az duyarlı bir versiyon önermişlerdir. Gerds ve arkadaşları ise ters olasılık sansürleme ağırlıklandırması (IPCW) tabanlı zamana bağlı bir C-index formülasyonu geliştirmişlerdir. Pratikte C-index değerlerinin yorumlanmasında şu genel

kılavuz kullanılmaktadır: 0.50-0.59 zayıf, 0.60-0.69 orta, 0.70-0.79 iyi, 0.80-0.89 çok iyi ve 0.90 üzeri mükemmel ayırt edicilik olarak değerlendirilmektedir (95).

2.5.2. Bütünleşik Brier Skoru (Integrated Brier Score)

Brier score, bir modelin belirli bir zaman noktasında tahmin ettiği sağkalım olasılığının gerçek durumdan sapmasını ölçen bir kalibrasyon metriğidir. Graf ve arkadaşları (94) sansürlü veriler için ters olasılık sansürleme ağırlıklandırması (IPCW) ile Brier score hesaplamasını önermiştir. Brier score formülü Denklem (2.9)'da verildi. Bu formülde $Y_i(t)$ bireyin t zamanındaki durumunu, $\hat{S}(t|X_i)$ modelin tahminini ve $w_i(t)$ IPCW ağırlığını temsil etmektedir.

$$BS(t) = \frac{1}{n} \sum_{i=1}^n w_i(t) \cdot \left(Y_i(t) - \hat{S}(t|X_i) \right)^2 \quad (2.9)$$

Bütünleşik Brier skoru (Integrated Brier Score, IBS), Brier score'un tüm zaman noktaları üzerindeki integralinin alınmasıyla elde edilmektedir. Daha düşük IBS değerleri daha iyi model performansını göstermektedir (98). IBS, hem ayırt edicilik hem de kalibrasyon bilgisini birleştiren kapsamlı bir metrik olarak değerlendirilmektedir (Denklem 2.10). Temel sınırlılığı, IPCW ağırlıklarının sansürleme mekanizmasının doğru modellenmesine bağlı olmasıdır. IBS değerinin referans modelle (Kaplan-Meier tahmincisi gibi) karşılaştırılması, modelin katma değerinin değerlendirilmesinde yararlıdır.

$$IBS = \frac{1}{t_{\max} - t_{\min}} \int_{t_{\min}}^{t_{\max}} BS(t) dt \quad (2.10)$$

Brier score'un yorumlanmasında referans modelle karşılaştırma önemlidir. Kaplan-Meier tahmincisi referans model olarak kullanıldığında, incelenen modelin IBS değerinin KM tahmincisinin IBS değerinden düşük olması modelin katma değer sağladığını göstermektedir. IBS Scaled = $1 - (IBS_{\text{model}} / IBS_{\text{KM}})$ formülü ile elde edilen ölçeklendirilmiş IBS değeri, R^2 istatistiğine benzer şekilde 0 ile 1 arasında yorumlanabilir bir metrik sunmaktadır (98). Bu metrik, modelin Kaplan-Meier tahminlerini ne ölçüde iyileştirdiğini göstermektedir.

Zamana bağlı AUC (Time-Dependent Area Under the ROC Curve), sağkalım modellerinin belirli zaman noktalarındaki ayırt edici performansını değerlendiren bir metriktir. Uno ve arkadaşları (99) tarafından önerilen kümülatif/dinamik AUC formülasyonunda, belirli bir zaman noktası t için t 'den önce olay yaşayan bireyler ile t 'den sonra hâlâ risk altında olan bireyler arasındaki ayırt edicilik ölçülmektedir.

Bu formülasyon sansürlemeye karşı daha dirençlidir ve bootstrap güven aralıkları ile değerlendirilebilmektedir (100). Zamana bağlı AUC'nin avantajı, modelin performansının farklı zaman noktalarında nasıl değiştiğini göstermesidir; bu sayede modelin erken ve geç dönem tahmin performansı ayrı ayrı değerlendirilebilmektedir. Genellikle modellerin erken dönemde daha yüksek AUC değerleri elde ettiği, takip süresi uzadıkça performansın azaldığı gözlenmektedir.

2.6. Çapraz Doğrulama ve Hiperparametre Optimizasyonu

Çapraz doğrulama (cross-validation), modelin genelleme performansını yansız bir şekilde tahmin etmek için kullanılan bir yeniden örnekleme tekniğidir. k-katlı çapraz doğrulamada veri k eşit parçaya bölünmekte, her iterasyonda k-1 parça eğitim, 1 parça test seti olarak kullanılmakta ve sonuçlar ortalanmaktadır (101). Sağkalım analizinde çapraz doğrulama uygulanırken sansürleme oranlarının katlar arasında dengeli dağılmasına dikkat edilmesi gerekmektedir (102).

İç içe çapraz doğrulama (nested cross-validation), hiperparametre optimizasyonu ve model değerlendirmesinin yansız bir şekilde yapılmasını sağlayan iki aşamalı bir yöntemdir. Dış döngü model performansını değerlendirirken, iç döngü hiperparametre ayarlamasını gerçekleştirmektedir. Varma ve Simon (22) iç içe çapraz doğrulamanın tek döngülü çapraz doğrulamaya kıyasla daha yansız performans tahminleri ürettiğini göstermiştir. Bu tez çalışmasında 5×3 iç içe çapraz doğrulama uygulanmıştır; dış döngüde 5 kat ile model performansı, iç döngüde 3 kat ile hiperparametre optimizasyonu gerçekleştirilmiştir (46, 103).

Hiperparametre optimizasyonu, model performansını belirleyen hiperparametrelerin en uygun değerlerini bulmayı amaçlamaktadır. Temel yaklaşımlar arasında ızgara araması (grid search), rastgele arama (random search) ve Bayesian optimizasyon yöntemleri yer almaktadır (22). Bayesian optimizasyon yöntemlerinden ağaç yapılı Parzen tahmincisi (Tree-structured Parzen Estimator, TPE), önceki denemelerin sonuçlarını kullanarak parametre uzayını akıllıca daraltmaktadır. Akiba ve arkadaşları (103) tarafından geliştirilen Optuna kütüphanesi, TPE tabanlı verimli hiperparametre optimizasyonu sunmaktadır. Bu tez çalışmasında parametre uzayı küçük olan modellerde ızgara araması, diğer modellerde Optuna TPE kullanıldı.

Makine öğrenmesi modellerinin sağkalım analizinde uygulanmasında dikkat edilmesi gereken önemli pratik konular bulunmaktadır. Veri sızıntısı (data leakage) önlenmesi amacıyla tüm ön işleme adımlarının (ölçeklendirme, kodlama, öznitelik

seçimi) çapraz doğrulama katı içinde gerçekleştirilmesi zorunludur; aksi takdirde performans tahminleri iyimser yönde yanlı olmaktadır (23, 50). Bu tez çalışmasında Z-standardizasyon her bir fold içinde ayrı ayrı uygulandı. Ayrıca kategorik değişkenlerin kodlanmasında one-hot encoding yöntemi kullanılmış ve bu işlem de fold içinde gerçekleştirildi. Nominal yapıdaki rx değişkeni (Obs, Lev, Lev+5FU) arasında doğal bir sıralama bulunmadığı için ordinal kodlama yerine one-hot encoding tercih edilmiştir; bu sayede her bir tedavi kolunun model çıktısına katkısı SHAP ve SurvLIME analizlerinde ayrı ayrı yorumlanabilmiştir. Python ortamında uygulanan SHAP ve SurvSHAP(t) analizlerinde sürekli, binary ve ordinal değişkenler orijinal sayısal halleriyle modele girdi; sadece nominal rx değişkeni one-hot encoding ile iki gösterge değişkene açıldı (Levamisol ve Lev+5FU; referans kategori Gözlem olarak düşürülmüştür). SurvLIME analizleri ise R survex paketi ile uygulanırken paketin factor mekanizması gereği tüm kategorik değişkenler otomatik olarak gösterge (dummy) değişkenlere açılmıştır; Tablo 4.6 ve Tablo 4.7'de raporlanan öznitelik adları (ör. Yayılım: Bitişik yapılar, Diferansiyasyon: Kötü, Tedavi: Lev+5FU) bu otomatik dönüşümün Türkçeleştirilmiş karşılıklarıdır. Lenf nodu tutulumu yalnızca sürekli pozitif lenf nodu sayısı (nodes) değişkeniyle modellenmiş, ayrıca bir node4 gösterge değişkeni kullanılmamıştır.

Model seçiminde performans metriklerinin yanı sıra hesaplama maliyeti, yorumlanabilirlik, klinik uygulanabilirlik ve model kararlılığı gibi çok boyutlu kriterlerin dikkate alınması gerekmektedir. En yüksek C-index değerine sahip model her zaman en iyi klinik model olmayabilir; basit ve yorumlanabilir bir modelin marjinal olarak düşük performansla klinik pratikte daha değerli olabileceği durumlar mevcuttur (44, 53). Bu nedenle tez çalışmasında performans karşılaştırmasının yanı sıra açıklanabilirlik boyutunun da sistematik olarak değerlendirilmesi, bütüncül bir model seçim çerçevesi sunmayı amaçlamaktadır.

Bootstrap yeniden örnekleme, model performansının güvenilirlik değerlendirmesinde yaygın olarak kullanılan bir istatistiksel yöntemdir. Efron ve Tibshirani (104) tarafından geliştirilen bu yaklaşımda, orijinal veri setinden yerine koymalı örnekleme ile B adet bootstrap örnekleme oluşturulmakta ve her örneklemede model eğitilerek performans metriği hesaplanmaktadır. B tekrarlı bootstrap dağılımından güven aralıkları elde edilmektedir. Bu tez çalışmasında B=1000 tekrarlı bootstrap ile %95 güven aralıkları hesaplandı.

2.7. İlgili Çalışmalar

Sağkalım analizinde makine öğrenmesi ve derin öğrenme yöntemlerinin karşılaştırılması, son yıllarda yoğun araştırma konusu olmuştur. Bu bölümde tez konusuyla doğrudan ilişkili olan önceki çalışmalar ve elde ettikleri temel bulgular özetlenmektedir.

Wiegerebe ve arkadaşları 2024 yılında derin öğrenme tabanlı sağkalım analizinin kapsamlı bir derlemesini yayımlamışlardır (13). Bu derlemede Cox tabanlı, ayırık zamanlı, sürekli zamanlı, yarışan riskli ve çoklu görevli derin öğrenme sağkalım modelleri sistematik olarak sınıflandırılmıştır. Yazarlar, derin öğrenme modellerinin geleneksel yöntemlere göre avantajının büyük ölçüde veri setinin boyutuna, öznelilik uzayının karmaşıklığına ve hedef ilişkinin doğrusallığına bağlı olduğunu vurgulamışlardır. Küçük tablo verilerinde ensemble ağaç modellerinin genellikle daha güvenilir sonuçlar ürettiği belirtilmiştir.

Sağkalım modellerinin klinik pratikte uygulanabilirliği üzerine de çeşitli çalışmalar yapılmıştır. Haider ve arkadaşları (24) bireysel sağkalım dağılımlarının oluşturulması ve değerlendirilmesi üzerine kapsamlı bir çalışma sunmuşlardır. Bu çalışmada model performansının değerlendirilmesinde yalnızca ayırt edicilik metriklerinin (C-index) yeterli olmadığı, kalibrasyon (IBS) ve bireysel tahmin doğruluğunun birlikte değerlendirilmesinin gerektiği vurgulanmıştır. Bu yaklaşım, tez çalışmasında hem C-index hem de IBS metriklerinin birlikte kullanılmasının teorik gerekçesini desteklemektedir.

Wang ve arkadaşları 2019 yılında yayımladıkları kapsamlı bir derleme çalışmasında, makine öğrenmesi tabanlı sağkalım analizi yöntemlerini sistematik olarak sınıflandırmışlardır (14). Çalışmada sağkalım analizine uyarlanmış makine öğrenmesi modellerinin kayıp fonksiyonları, çıktı türleri ve uygulama alanları detaylı şekilde tartışılmıştır. Sonuçlar, ensemble yöntemlerin çoğu veri setinde Cox modeline kıyasla marjinal ancak tutarlı bir performans artışı sağladığını gösterdi. Bununla birlikte, derin öğrenme modellerinin küçük ve orta ölçekli tablo verilerde beklenen kadar üstünlük göstermediği rapor edildi.

Kvamme ve arkadaşları (76) 2019 yılında DeepSurv, DeepHit ve diğer derin öğrenme sağkalım modellerini sistematik olarak karşılaştırmışlardır. Çalışma, derin öğrenme modellerinin Cox modeline göre avantajının veri setinin boyutuna ve değişkenler arasındaki ilişkinin karmaşıklığına bağlı olduğunu gösterdi. Küçük veri

setlerinde Cox modelinin, büyük ve karmaşık veri setlerinde derin öğrenme modellerinin daha üstün performans sergilediği bulunmuştur. Ayrıca, sürekli zamanlı modellerin (DeepSurv) ayırık zamanlı modellere (DeepHit) kıyasla daha düşük varyans gösterdiği raporlanmıştır.

Spooner ve arkadaşları (31, 105) 2020 yılında makine öğrenmesi tabanlı sağkalım modellerinin kapsamlı bir karşılaştırmasını sunmuşlardır. Çalışmada RSF, gradient boosting, SVM ve derin öğrenme modelleri 30 farklı veri setinde değerlendirilmiştir. RSF'nin genel olarak en tutarlı performansı sergilediği, XGBoost'un bazı veri setlerinde üstün sonuçlar verdiği ve tek bir modelin tüm veri setlerinde dominant olmadığı rapor edildi. Bu bulgu, no free lunch teoremiyle uyumlu olarak farklı veri yapılarının farklı modelleri desteklediğini göstermektedir (48).

Christodoulou ve arkadaşları 2019 yılında yayımladıkları sistematik derlemede, klinik tahmin modelleri için makine öğrenmesinin lojistik regresyona karşı performans avantajını değerlendirmişlerdir (53). 71 çalışmanın meta-analizi sonucunda makine öğrenmesinin istatistiksel olarak anlamlı ancak klinik olarak marjinal bir performans artışı sağladığı gösterildi. Bu bulgu, model karmaşıklığının artırılmasının her zaman pratik fayda sağlamadığını ve açıklanabilirliğin performans kadar önemli olduğunu vurgulamaktadır.

Açıklanabilir sağkalım analizi alanında Krzyżiński ve arkadaşları SurvSHAP(t) yöntemini tanıtarak zamana bağlı açıklamaların önemini vurgulamışlardır (18). Yöntem, standart SHAP değerlerinin zamana bağlı sağkalım fonksiyonlarına genellemesi olarak formüle edilmiş ve survex R paketinde uygulanmıştır (91). Biecek ve Burzykowski (20) survex paketini kullanarak sağkalım modellerinin açıklanabilirliği üzerine kapsamlı bir çerçeve sunmuşlardır.

Kovalev ve arkadaşları (19) SurvLIME yöntemini önererek sağkalım modellerinin bireysel düzeyde açıklanabilirliğini sağlamışlardır. Yöntem, Cox katsayıları cinsinden lokal açıklamalar üreterek klinisyenler için yorumlanabilir çıktılar sunmaktadır. Pachón-García ve arkadaşları SurvLIME (101)'in Python uygulamasını (SurvLIMEpy) yayımlayarak yöntemin erişilebilirliğini artırmışlardır. Bu çalışmada farklı kernel fonksiyonları ve pertürbasyon stratejileri ile SurvLIME performansı değerlendirilmiştir.

Moncada-Torres ve arkadaşları (106) 2021 yılında meme kanseri sağkalım verilerinde açıklanabilir makine öğrenmesinin Cox regresyon tahminlerinden üstün performans gösterebileceğini ve klinik olarak anlamlı bilgiler sunabileceğini göstermişlerdir. Kolon kanseri özelinde Adatorwovor ve arkadaşları (107) makine

öğrenmesi modellerini Kentucky Kanser Kayıt (KCR) veri tabanı üzerinde değerlendirmişlerdir. Rastgele sağkalım ormanı ve Lasso modellerinin geleneksel Cox yaklaşımına kıyasla daha yüksek tahmin doğruluğu (C-index) sağladığı gösterilmiştir; ancak çalışmada açıklanabilirlik boyutu ele alınmadı.

Sağkalım analizinde açıklanabilirlik yöntemlerinin karşılaştırılmasına odaklanan çalışmalar henüz sınırlıdır. Utkin ve arkadaşları (85) SurvNAM (Survival Neural Additive Model) yaklaşımını önererek sinir ağlarına dayalı yorumlanabilir sağkalım modelleri geliştirmişlerdir. Bu yaklaşımda her bir özneliliğin etkisi ayrı bir alt ağ tarafından modellenmekte ve bireysel etkilerin toplamı olarak genel tahmin elde edilmektedir. Bununla birlikte, SurvNAM değişkenler arası etkileşimleri yakalayamamaktadır.

Türkiye'de sağkalım analizinde makine öğrenmesi uygulamalarına ilişkin çalışmalar da mevcuttur. Biyoistatistik alanında yapılan çalışmalarda genellikle RSF ve Cox modeli karşılaştırılmakta, ancak derin öğrenme modelleri ve açıklanabilirlik yöntemleri nadiren ele alınmaktadır (113). Bu tez çalışması, on farklı modelin kapsamlı karşılaştırmasını ve üç açıklanabilirlik yönteminin sistematik uygulamasını içermesiyle mevcut ulusal literatüre özgün bir katkı sağlamayı hedeflemektedir.

Steingrimsson ve arkadaşları (112) sansürleme yansızlığı düzeltilmiş sağkalım ağaçları ve ensemblesini önererek sansürleme yanlılığının model performansı üzerindeki etkisini azaltmışlardır. Bu yaklaşımda ters olasılık sansürleme ağırlıkları bölme kriterine entegre edilmekte ve sansürleme mekanizmasının bilgilendirici olabileceği durumlarda daha güvenilir tahminler üretilmektedir. Sonabend ve arkadaşları (25) mlr3proba paketini geliştirerek R ekosisteminde sağkalım modelleri için standartlaştırılmış bir karşılaştırma çerçevesi sunmuşlardır.

Makine öğrenmesi sağkalım modellerinin farmakovijilans, klinik deneyler ve sağlık ekonomisi gibi alanlarda da uygulamaları artmaktadır. Klinik deneylerde tedavi etkinliğinin değerlendirilmesinde heterojen tedavi etkilerinin (heterogeneous treatment effects) tahmini, bireyselleştirilmiş tıp yaklaşımının temelini oluşturmaktadır. DeepSurv modelinin tedavi öneri sistemi olarak kullanılabilmesi bu bağlamda önem taşımaktadır (11). Sağlık ekonomisinde ise maliyet-etkililik analizleri için doğru sağkalım tahminleri gereklidir ve makine öğrenmesi modelleri bu alanda giderek daha fazla kullanılmaktadır.

Sonuç olarak, mevcut literatürdeki boşluklar şu şekilde özetlenebilir: (i) çok sayıda model ailesinin (istatistiksel, ağaç tabanlı, ensemble, kernel, derin öğrenme) aynı çalışmada sistematik karşılaştırılmasının sınırlı olması, (ii) performans

değerlendirmesinin yalnızca C-index ile değil, IBS ve TD-AUC gibi tamamlayıcı metriklerle birlikte yapılmasının nadir olması, (iii) açıklanabilirlik yöntemlerinin (SHAP, SurvSHAP(t), SurvLIME) birlikte uygulanıp karşılaştırıldığı çalışmaların bulunmaması, (iv) iç içe çapraz doğrulama ve bootstrap güven aralıkları ile desteklenen metodolojik olarak sağlam değerlendirmelerin yetersiz olması. Bu tez çalışması, tüm bu boyutları tek bir çerçevede ele alarak mevcut boşlukları kapatmayı amaçlamaktadır.

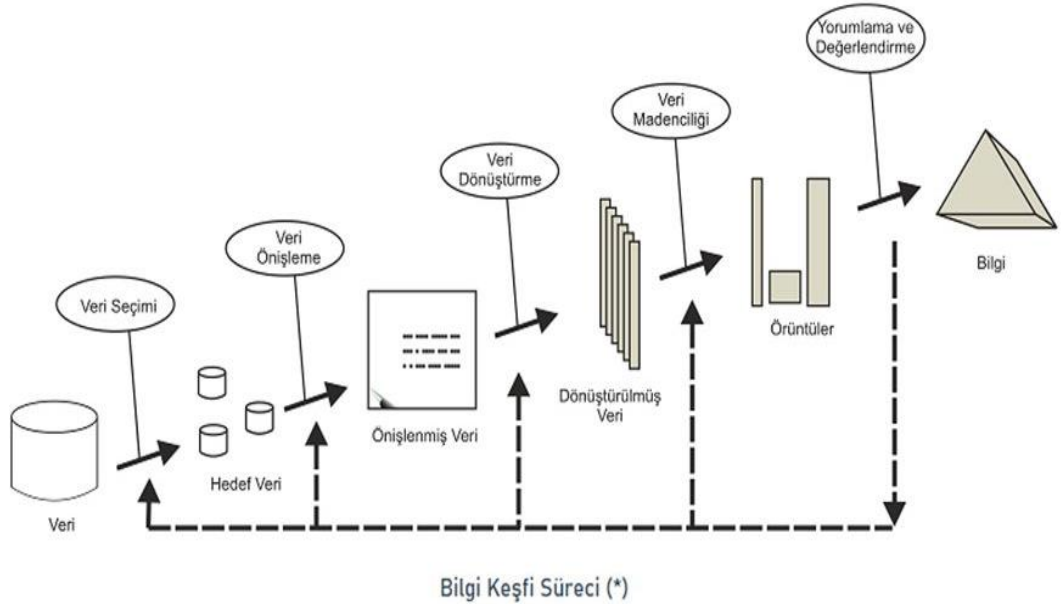
Ayrıca mevcut literatür, makine öğrenmesi ve derin öğrenme yöntemlerinin sağkalım analizinde güçlü bir potansiyel taşımakla birlikte, tek bir modelin tüm durumlarda üstün olmadığını ortaya koymaktadır (114, 115).

Literatürdeki çalışmaların bir diğer ortak sınırlılığı, açıklanabilirlik yöntemleri arasındaki tutarlılığın yeterince araştırılmamış olmasıdır. SHAP, SurvSHAP(t) ve SurvLIME farklı teorik temellere dayanmakta ve farklı türde açıklamalar üretmektedir (116). Bu yöntemlerin aynı veri seti ve aynı modeller üzerinde sistematik karşılaştırılması, her bir yöntemin güçlü ve zayıf yönlerinin ortaya konması açısından değerli bilgiler sunacaktır. Bu tez çalışması, bu karşılaştırmalı perspektifi sağlamayı hedeflemektedir.

3. MATERYAL VE METOT

3.1. Veri Seti

Çalışmada "survival" R paketinde yer alan colon veri seti kullanıldı (42). Veri seti, kolon kanseri tedavisinde levamisol ve 5-florourasil kemoterapisinin etkinliğini değerlendiren randomize kontrollü bir klinik çalışmadan elde edildi (108). Orijinal veri seti, 929 hastaya ait iki olay tipini (nüks ve ölüm) içeren toplam 1858 kayıttan oluşmaktaydı; her hasta için biri nüks (etype=1), diğeri ölüm (etype=2) olmak üzere iki satır yer almaktaydı. Genel sağkalım (overall survival) analizi amaçlandığından yalnızca ölüm olayına (etype=2) ait 929 hasta seçildi. Eksik değer içeren 41 hasta listwise silme yöntemiyle çıkarıldı. Ardından Tukey-IQR yöntemiyle aykırı değer analizi yapıldı; ağaç tabanlı modellerin aykırı değerlere dayanıklılığı ve klinik bilgi kaybını önleme gerekçesiyle aykırı gözlemler veri setinde tutuldu ve nihai örneklem büyüklüğü $n=888$ hastaya ulaştı. Analizler, nodes (pozitif lenf düğümü sayısı) ile yüksek çoklu bağlantı gösteren node4 (pozitif lenf düğümü sayısının dörtten fazla olup olmadığını gösteren ikili değişken) modellemeyi dışlandıktan sonra 10 bağımsız değişken üzerinden yürütüldü. Değişken tanımları Tablo 3.1’de, analiz akışı Şekil 3.1’de verildi.



Şekil 3.1. Bilgi keşfi süreci

Şekil 3.1’de gösterilen veri madenciliği akışı, Fayyad ve arkadaşlarının 1996 yılında önerdiği Bilgi Keşfi Süreci (Knowledge Discovery in Databases, KDD)

çerçevesini temel aldı (119). KDD; ham veriden anlamlı, geçerli, yeni ve potansiyel olarak yararlı örüntülerin sistematik biçimde elde edilmesi için tanımlanmış beş ardışık aşamadan oluşur: veri seçimi, önışleme, dönüştürme, veri madenciliği ve yorumlama-değerlendirme. Her aşama, bir önceki adımın çıktısını girdi olarak kullanır ve sürecin sonunda elde edilen örüntüler, alan uzmanı yorumlamasıyla bilgiye dönüşür. Bu çalışmada da analitik akış, her bir KDD aşamasının somut metodolojik karşılıkları belirlenerek tasarlandı.

Dönüştürme aşamasında nominal rx (tedavi yöntemi) değişkeni one-hot encoding ile kodlanırken sayısal değişkenler (age ve nodes) her çapraz doğrulama katında ayrı ayrı Z-standardizasyona tabi tutuldu; böylece veri sızıntısı (data leakage) riski önleildi. Veri madenciliği aşamasında klasik istatistiksel, ağaç tabanlı ensemble, uzaklık tabanlı, kernel tabanlı ve derin öğrenme yaklaşımlarını temsil eden on farklı sağkalım modeli karşılaştırmalı olarak eğitildi. Son aşama olan yorumlama-değerlendirmede SHAP, SurvSHAP(t) ve SurvLIME olmak üzere üç farklı XAI yöntemi katmanlı bir yaklaşımla uygulanarak modellerin tahminleri yorumlanabilir hale getirildi. Bu aşamaların çalışmadaki somut karşılıkları Tablo 3.1.'de özetlendi.

Tablo 3.1. KDD aşamalarının bu çalışmadaki karşılıkları

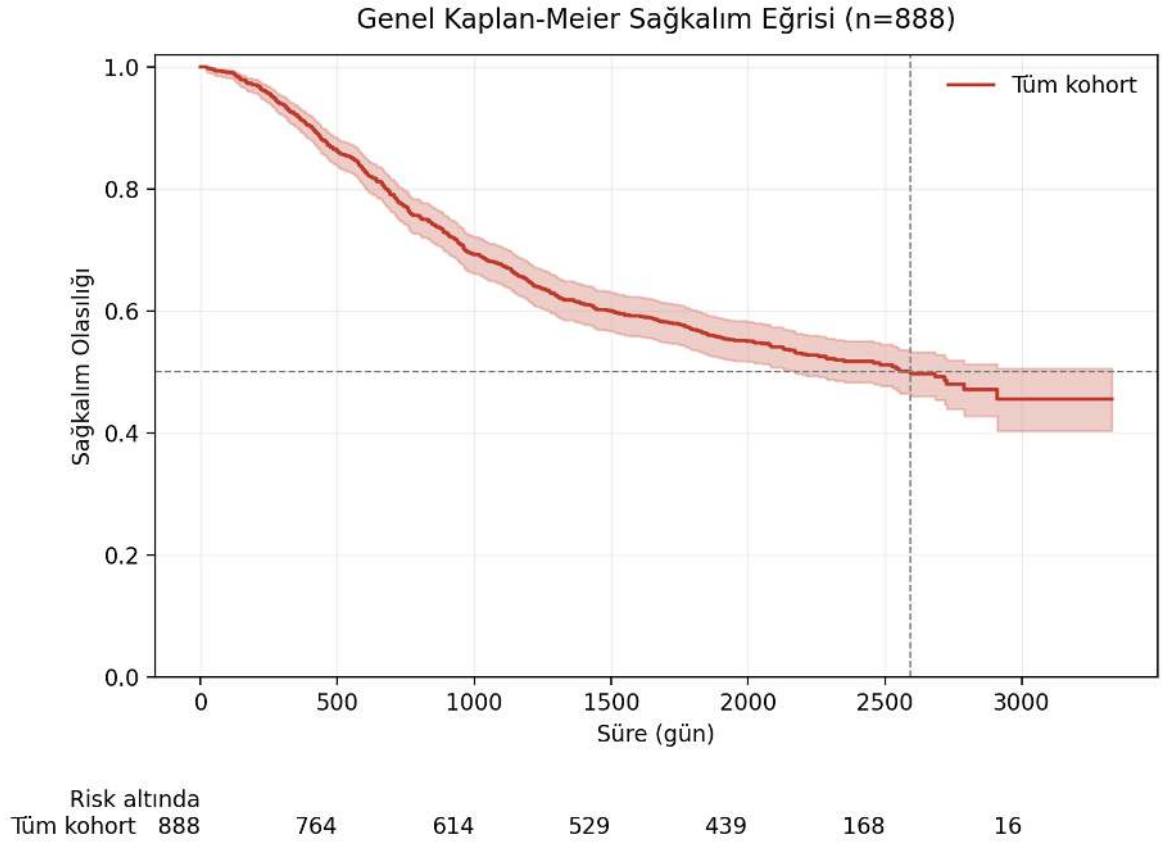
KDD Aşaması	Bu Çalışmadaki Karşılığı
Veri Seçimi	R survival paketi colon veri seti; 1858 kayıt → 929 hasta (etype=2 alt grubu) → 888 hasta (listwise sonrası)
Önişleme	Listwise silme (41 satır), Tukey-IQR aykırı değer analizi (75 gözlem tutuldu); n = 888
Dönüştürme	rx → one-hot encoding (referans: Gözlem; Levamisol ve Lev+5FU gösterge değişkenleri); age ve nodes → fold-içi Z-standardizasyon (data leakage önlemi)
Veri Madenciliği	10 sağkalım modeli: CoxPH, CART, RSF, ExtraSurvTrees, GBSurv, XGBoost, KNN, SurvivalSVM, DeepSurv, DeepHit
Yorumlama-Değerlendirme	XAI katmanları: SHAP (global statik), SurvSHAP(t) (global zamana bağlı), SurvLIME (lokal bireysel)

Veri seti yedi kategorik (rx, sex, obstruct, perfor, adhere, surg ve node4), iki sıralı kategorik (differ ve extent), bir kesikli sayım türündeki (nodes) ve iki sürekli (age ve time) değişkenden oluşmaktadır. Sağkalım analizlerinde bağımlı değişkenler olarak takip süresini temsil eden time ve olay durumunu temsil eden status değişkenleri kullanıldı. Model geliştirme aşamasında node4 değişkeninin, pozitif lenf nodu sayısını ifade eden nodes değişkeninden türetilmiş olması ve iki değişken arasında yüksek düzeyde çoklu bağlantı bulunması nedeniyle bağımsız değişkenler arasından çıkarılması uygun görüldü (Tablo 3.2).

Tablo 3.2. Veri setindeki değişkenlerin tanımları

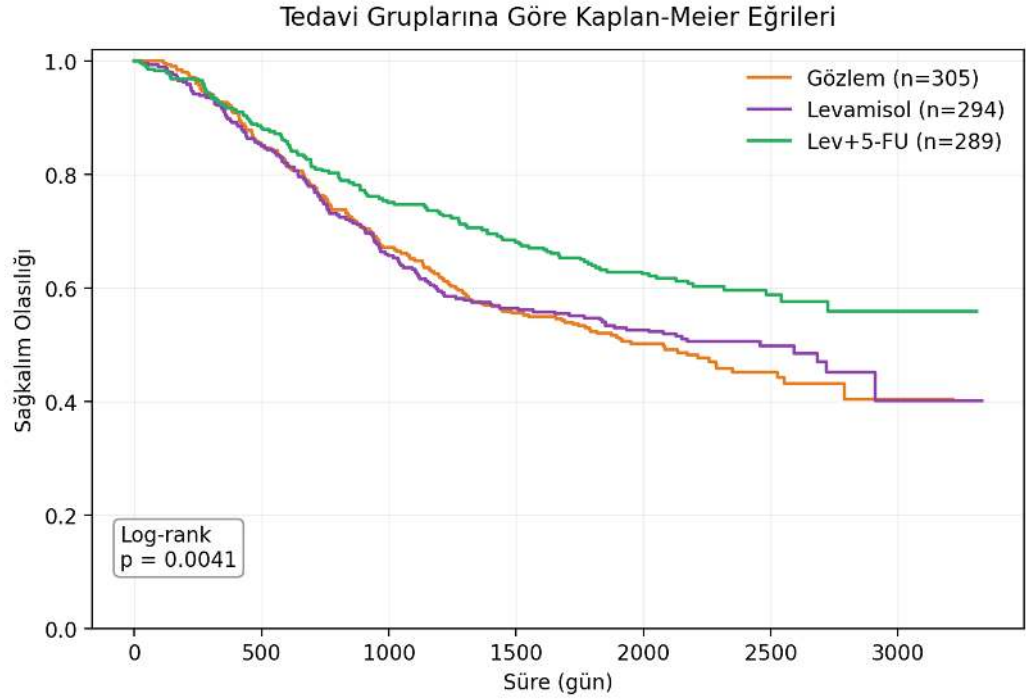
Değişken	Veri Türü	Kodlama	Açıklama
rx	Kategorik	1=Lev+5-FU, 2=Gözlem, 3=Levamisol	Tedavi grupları
sex	Kategorik	0=Kadın, 1=Erkek	Cinsiyet
age	Sürekli	Yıl	Hastanın yaşı
obstruct	Kategorik	0=Hayır, 1=Evet	Kalın bağırsak tıkanıklığı
perfor	Kategorik	0=Hayır, 1=Evet	Kalın bağırsak perforasyonu
adhere	Kategorik	0=Hayır, 1=Evet	Komşu organlara yapışma
nodes	Kesikli (sayım)	Pozitif lenf düğümü sayısı	Pozitif lenf nodu sayısı
differ	Sıralı kategorik	1=İyi, 2=Orta, 3=Kötü	Tümör diferansiyasyon derecesi
extent	Sıralı kategorik	1=Submukoza, 2=Kas tabakası, 3=Seroza, 4=Bitişik yapılar	Tümör yayılım düzeyi
surg	Kategorik	0=Kısa, 1=Uzun	Cerrahi ile randomizasyon arasındaki süre
node4	Kategorik	0=Hayır, 1=Evet	≥4 pozitif lenf düğümü varlığı
time	Sürekli	Gün	Takip süresi
status	Kategorik	0=Sansür, 1=Ölüm	Sağkalım olay değişkeni

Hedef değişkenler status ve time'dır. Sağkalım eğrileri Şekil 3.2 ve Şekil3.3'te verildi. Şekil 3.2'de 888 hastanın genel sağkalım eğrisi yer aldı. Genel medyan sağkalım süresi 2593 gün (yaklaşık 7,1 yıl) olarak hesaplandı (%95 GA alt sınır: 2171 gün; üst sınır izlem süresi yetersizliği nedeniyle tahmin edilemedi). Tahmin edilen sağkalım olasılıkları birinci yılda 0,914, ikinci yılda 0,776, üçüncü yılda 0,677 ve beşinci yılda 0,567 düzeyindeydi. Risk altındaki birey sayısı 1000. günde 614, 2000. günde 439, 3000. günde 17 idi. Olay hızının özellikle ilk 1000 günde yoğunlaştığı, sonrasında eğrinin eğiminin azaldığı gözlemlendi.



Şekil 3.2. Genel Kaplan-Meier sağkalım eğrisi

Şekil 3.3'te tedavi grupları arasında log-rank testiyle anlamlı fark gözlemlendi ($\chi^2=11,02$; $sd=2$; $p=0,0041$). Lev+5-FU grubu izlem boyunca en yüksek, Gözlem grubu en düşük sağkalım olasılığını gösterdi. Grup bazında medyan sağkalım süreleri Gözlem grubunda 2077 gün, Levamisol grubunda 2458 gün olarak hesaplanırken Lev+5-FU grubunda medyana izlem süresi içinde ulaşılmadı (%95 GA alt sınırı: 2725 gün). İkili karşılaştırmalar (Tablo 3.3), gruplar arası genel farkın tümüyle Lev+5-FU kombinasyonundan kaynaklandığını ortaya koydu: Lev+5-FU hem Gözlem grubundan ($p=0,0014$) hem de Levamisol grubundan ($p=0,0102$) anlamlı biçimde daha iyi sağkalım sağlarken, Gözlem ile Levamisol arasında anlamlı fark bulunmadı ($p=0,5844$). Üç karşılaştırma için Bonferroni düzeltmesi uygulandığında ($\alpha=0,0167$) Lev+5-FU içeren her iki karşılaştırma anlamlılığını korudu.



	Risk altında						
Gözlem (n=305)	305	259	204	169	137	49	5
Levamisol (n=294)	294	250	193	166	141	56	4
Lev+5-FU (n=289)	289	255	217	194	161	63	7

Şekil 3.3. Tedavi gruplarına göre Kaplan-Meier eğrileri

Tablo 3.3. Tedavi grupları arası ikili log-rank karşılaştırmaları

Karşılaştırma	χ^2	sd	p
Gözlem – Levamisol	0,299	1	0,5844
Gözlem – Lev+5-FU	10,253	1	0,0014
Levamisol – Lev+5-FU	6,597	1	0,0102

sd: serbestlik derecesi. Bonferroni düzeltmesi sonrası anlamlılık eşiği $\alpha=0,0167$.

3.1.1. Önışleme ve Dönüştürme

Kayıp değer yoktur. Aykırı değer analizi sonucunda 75 gözlem tespit edildi (97). Ağaç tabanlı modeller aykırı değerlere dayanıklı olduğundan ve klinik bilgi kaybını önlemek amacıyla bu gözlemler veri setinde tutuldu. rx (tedavi grubu) değişkeni one-hot encoding ile kodlandı. Z-standardizasyon yalnızca sayısal değişkenlere (yaş, pozitif lenf düğümü sayısı) fold içinde uygulandı (50). Ölçekleme stratejisi Tablo 3.4'de verildi. Tablo 3.4'e göre beş modelde ölçekleme uygulanmış olupbeşinde uygulanmadı.

Tablo 3.4. Ölçeklendirme stratejisi

Model	Ölçekleme	Gerekçe
CoxPH, KNN, SVM, DeepSurv, DeepHit	Evet	Katsayı/uzaklık/gradient duyarlılığı
RSF, CART, GBSurv, ExtraTrees, XGBoost	Hayır	Sıralama tabanlı dallanma

3.1.2. Modeller ve Optimizasyon

On sağkalım modeli kullanıldı (Tablo 3.3). Model seçiminde istatistiksel (6), ağaç tabanlı (57), ensemble (8, 66), uzaklık/kernel tabanlı (68, 72) ve derin öğrenme (11, 12) yaklaşımlarının temsil edilmesi hedeflendi. CoxPH referans model olarak, CART yorumlanabilirlik karşılaştırması için, RSF ve ExtraSurvTrees doğrusal olmayan ilişkilerin yakalanması için, XGBoost ve GradientBoosting güçlü düzenleme kapasitesi için, KNN ve SurvivalSVM parametrik olmayan alternatifler olarak, DeepSurv ve DeepHit ise derin öğrenmenin geleneksel modellere kıyasla performans artışı sağlayıp sağlayamadığını test etmek amacıyla seçildi.

Tablo 3.5'te beş modelin sağkalım fonksiyonu ürettiği, beşinin yalnızca risk skoru verdiği görüldü.

Tablo 3.5. Modeller

Model	Tip	Survival Fonk.
CoxPH	İstatistiksel	Evet
RSF	Ensemble	Evet
CART	Tek ağaç	Evet
GBSurv	Ensemble	Evet
ExtraTrees	Ensemble	Evet
XGBoost	Ensemble	Hayır
KNN	Uzaklık	Hayır
SVM	Kernel	Hayır
DeepSurv	Derin öğrenme	Hayır
DeepHit	Derin öğrenme	Hayır

Tablo 3.6'ya göre parametre uzayı küçük olan üç modelde (CoxPH, KNN, SurvivalSVM) exhaustive grid search, yedi modelde Optuna TPE Bayesian optimizasyonu kullanıldı (103, 80). TPE yöntemi önceki denemelerin sonuçlarına dayalı olarak parametre uzayını akıllıca daraltarak daha verimli arama yaptı.

Tablo 3.6. Optimizasyon yöntemi

Model	Yöntem	Deneme
CoxPH	Grid	6
KNN	Grid	24
SVM	Grid	30
RSF	Optuna TPE	50
CART	Optuna TPE	40
GBSurv	Optuna TPE	50
ExtraTrees	Optuna TPE	50
XGBoost	Optuna TPE	60
DeepSurv	Optuna TPE	50
DeepHit	Optuna TPE	60

3.1.3. Değerlendirme

Model performansının yansız tahmini için iç içe çapraz doğrulama (nested CV, 5×3) uygulandı (22). Performans metrikleri olarak Harrell'in C-index'i, Integrated Brier Score (94) ve kümülatif/dinamik AUC (99) kullanıldı. Model güvenilirliği 1000 tekrarlı bootstrap ile değerlendirildi (104). Açıklanabilirlik yöntemleri olarak SHAP, SurvSHAP(t) ve SurvLIME uygulandı (16, 18, 19). Sağkalım fonksiyonu üretmeyen modellerde Breslow baseline tahmini kullanıldı (47).

3.1.4. Yazılım Ortamı ve Kütüphaneler

Model eğitimi, hiperparametre optimizasyonu, SHAP ve SurvSHAP(t) analizleri Python 3.10 programlama dili ile yürütüldü. SurvLIME analizi ise R 4.3 programlama dili ve survex paketi kullanılarak gerçekleştirildi.

Python ortamında sağkalım modelleri scikit-survival kütüphanesi ile uygulandı. CoxPH, RSF, CART, GradientBoosting, ExtraSurvTrees ve SurvivalSVM modelleri bu kütüphaneden sağlandı. XGBoost sağkalım modeli xgboost kütüphanesinin Cox regresyon hedef fonksiyonu ile gerçekleştirildi. DeepSurv ve DeepHit derin öğrenme modelleri pycox kütüphanesi aracılığıyla PyTorch altyapısında eğitildi.

Hiperparametre optimizasyonunda Optuna kütüphanesinin TPE (Tree-structured Parzen Estimator) örnekleyicisi kullanıldı. Veri ön işleme ve çapraz doğrulama işlemleri scikit-learn ile gerçekleştirildi.

Açıklanabilirlik analizlerinde üç farklı kütüphane kullanıldı. Model-agnostik SHAP değerleri shap kütüphanesi ile hesaplandı. Zamana bağlı SurvSHAP(t) açıklamaları survshap Python paketi ile elde edildi. SurvLIME lokal açıklamaları ise R survex paketi aracılığıyla hesaplandı. R ortamında sağkalım modeli survival ve RSF modeli ranger paketleri ile oluşturuldu.

Grafik üretiminde Python tarafında matplotlib ve seaborn, R tarafında ggplot2 kütüphaneleri kullanıldı. Tüm analizlerde tekrarlanabilirliği sağlamak amacıyla rastgele tohum değeri 42 olarak sabitlendi.

4. BULGULAR

İç içe çapraz doğrulama sonuçları Tablo 4.1'de verildi. İç içe çapraz doğrulama sonuçlarına göre en yüksek ortalama C-index değerleri DeepSurv (0.6569 ± 0.0211), RSF (0.6568 ± 0.0188) ve DeepHit (0.6567 ± 0.0222) modellerinde birbirine çok yakın elde edildi; en düşük değer ise CART modelinde (0.6363 ± 0.0227) görüldü. Breslow baseline tahmini aracılığıyla tüm modellerde hesaplanan IBS değerlerinde RSF en düşük kalibrasyon hatasını (0.1743 ± 0.0099), XGBoost ise en yüksek değeri (0.1888 ± 0.0137) sergiledi. Zamana bağlı ayırt edicilik açısından SurvivalSVM (0.7162 ± 0.0288) ve RSF (0.7161 ± 0.0212) en yüksek TD-AUC değerlerine ulaştı.

Tablo 4.1. İç içe çapraz doğrulama (5×3) sonuçları

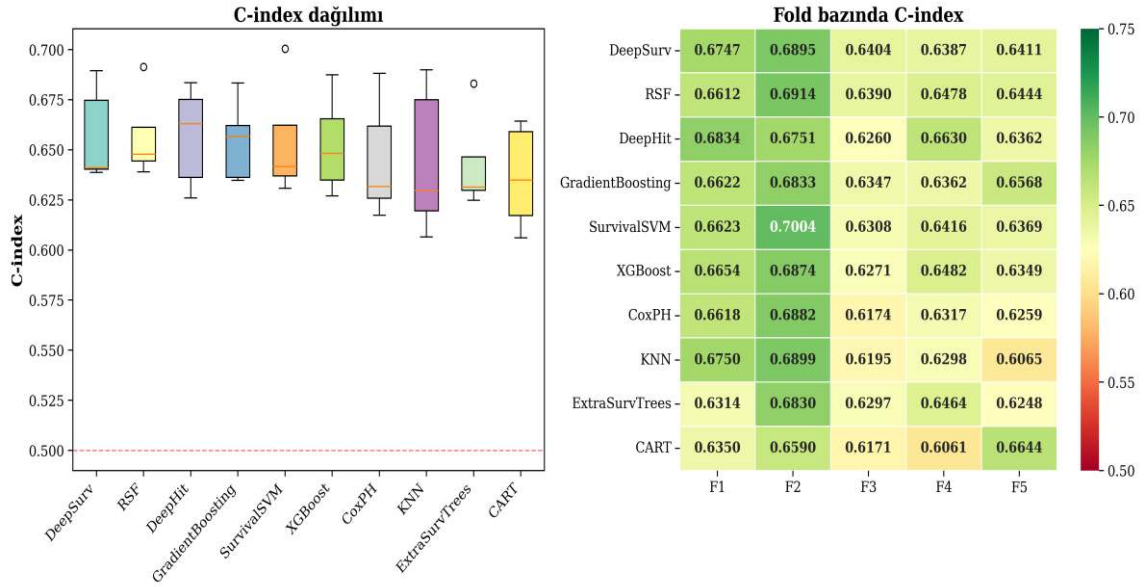
Sıra	Model	C-index (ort $\pm$ ss)	IBS (ort $\pm$ ss)	TD-AUC (ort $\pm$ ss)
1	DeepSurv	0.6569 ± 0.0211	0.1849 ± 0.0144	0.7151 ± 0.0222
2	RSF	0.6568 ± 0.0188	0.1743 ± 0.0099	0.7161 ± 0.0212
3	DeepHit	0.6567 ± 0.0222	0.1835 ± 0.0149	0.7157 ± 0.0213
4	GradientBoosting	0.6546 ± 0.0180	0.1755 ± 0.0090	0.7111 ± 0.0262
5	SurvivalSVM	0.6544 ± 0.0253	0.1839 ± 0.0123	0.7162 ± 0.0288
6	XGBoost	0.6526 ± 0.0217	0.1888 ± 0.0137	0.7102 ± 0.0252
7	CoxPH	0.6450 ± 0.0263	0.1781 ± 0.0110	0.7022 ± 0.0331
8	KNN	0.6441 ± 0.0325	0.1872 ± 0.0163	0.6954 ± 0.0385
9	ExtraSurvTrees	0.6431 ± 0.0212	0.1806 ± 0.0088	0.7028 ± 0.0291
10	CART	0.6363 ± 0.0227	0.1792 ± 0.0053	0.6790 ± 0.0287

Model güvenilirliğini değerlendirmek amacıyla 1000 tekrarlı bootstrap ile %95 güven aralıkları hesaplandı. Sonuçlar Tablo 4.2'de verildi. Tablo 4.2'ye göre bootstrap analizinde ExtraSurvTrees (0.7350) ve DeepSurv (0.7278) en yüksek C-index değerlerine sahiptir. Ancak ExtraSurvTrees iç içe çapraz doğrulamada 9. sırada yer aldığından, bu yüksek bootstrap değeri tüm veriye aşırı uyumdan kaynaklanan iyimser yanlılığı yansıttı.

Tablo 4.2. Bootstrap %95 güven aralıkları (B=1000)

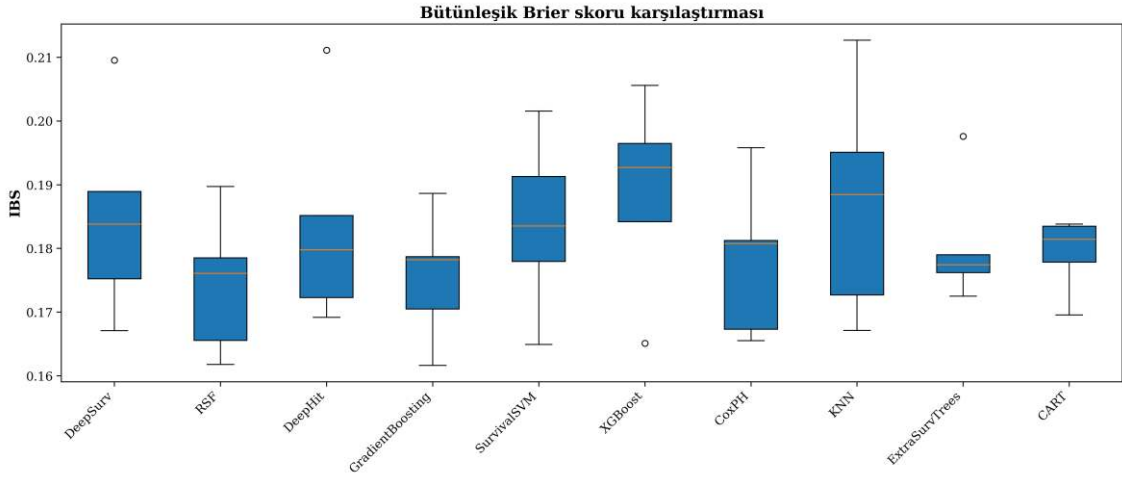
Sıra	Model	C-index [%95 CI]	IBS [%95 CI]	TD-AUC [%95 CI]
1	ExtraSurvTrees	0.7350 [0.7141, 0.7577]	0.1503 [0.1433, 0.1575]	0.8098 [0.7839, 0.8371]
2	DeepSurv	0.7278 [0.7064, 0.7512]	0.1538 [0.1440, 0.1630]	0.7901 [0.7622, 0.8215]
3	XGBoost	0.7178 [0.6961, 0.7426]	0.1598 [0.1489, 0.1696]	0.7826 [0.7542, 0.8139]
4	RSF	0.7175 [0.6958, 0.7416]	0.1586 [0.1512, 0.1659]	0.7871 [0.7581, 0.8173]
5	GradientBoosting	0.7154 [0.6929, 0.7410]	0.1553 [0.1471, 0.1631]	0.7784 [0.7495, 0.8113]
6	CART	0.7060 [0.6832, 0.7298]	0.1570 [0.1478, 0.1666]	0.7641 [0.7345, 0.7947]
7	DeepHit	0.6937 [0.6700, 0.7185]	0.1701 [0.1582, 0.1813]	0.7422 [0.7101, 0.7776]
8	SurvivalSVM	0.6720 [0.6477, 0.6971]	0.1763 [0.1654, 0.1871]	0.7370 [0.7038, 0.7686]
9	KNN	0.6676 [0.6414, 0.6926]	0.1779 [0.1671, 0.1882]	0.7233 [0.6898, 0.7566]
10	CoxPH	0.6673 [0.6429, 0.6926]	0.1726 [0.1647, 0.1809]	0.7270 [0.6936, 0.7592]

Şekil 4.1'in sol panelinde en yüksek medyan DeepHit (≈ 0.663), en düşük medyan ise KNN (≈ 0.630) modelinde görüldü; en geniş fold-içi dağılım KNN ve CART modellerinde gözlemlendi. DeepSurv en yüksek ortalamaya sahip olmasına karşın medyanı (≈ 0.641) orta düzeydedir; bu durum, dağılımının birinci ve ikinci foldlardaki yüksek değerlerce (0.6747 ve 0.6895) sağa çarpık hâle gelmesinden kaynaklandı. Sağ panelde ikinci fold (F2) modellerin çoğunda en yüksek C-index değerini verdi (SurvivalSVM 0.7004; RSF 0.6914; KNN 0.6899). DeepSurv dâhil tüm model–fold kombinasyonları için C-index hesaplandı (DeepSurv F4 = 0.6387).



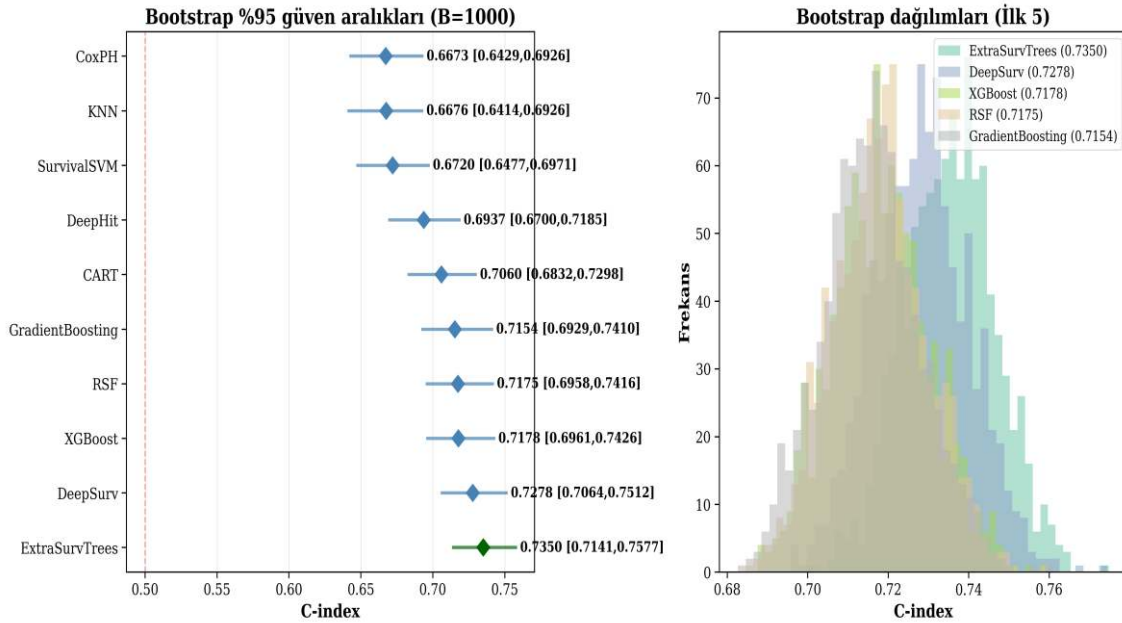
Şekil 4.1. C-index kutu grafiği ve fold bazında ısı haritası

Şekil 4.2'de Breslow baseline tahmini ile tüm 10 modelde hesaplanan IBS değerlerinin boxplot karşılaştırması sunuldu. RSF en düşük medyan IBS değerine (0.1741), XGBoost en yüksek medyan IBS değerine (0.1924) sahiptir. İlk dört sırada yer alan RSF, ExtraSurvTrees, CoxPH ve GradientBoosting modelleri doğrudan sağkalım fonksiyonu üreten modellerdir. Breslow wrapper ile hesaplanan modellerde (DeepSurv, SurvivalSVM, KNN, DeepHit, XGBoost) IBS değerleri genel olarak daha yüksektir.



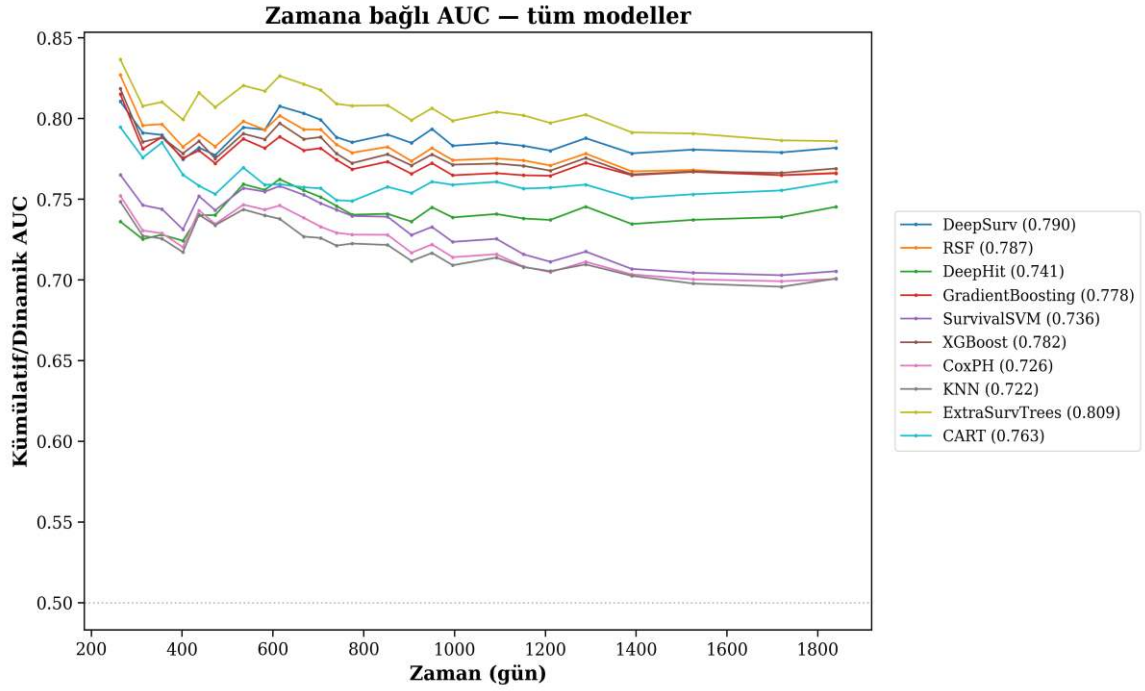
Şekil 4.2. Bütünleşik Brier skoru karşılaştırması

Şekil 4.3'nin sol panelinde RSF ve ExtraSurvTrees güven aralıkları diğer modellerden sağda konumlandı. CoxPH, KNN ve SurvivalSVM güven aralıkları birbirleriyle örtüştü. Sağ panelde bootstrap dağılımları yaklaşık normal dağılım gösterdi.



Şekil 4.3. Orman grafiği ve bootstrap dağılımları

Şekil 4.4'te ExtraSurvTrees (ortalama 0.789) ve RSF (0.785) en yüksek değerleri aldı. Tüm modellerde AUC ilk 600 günde 0.75-0.82 aralığında, 1000 günden sonra 0.70-0.77 aralığına düştü. SurvivalSVM 1200 günden sonra en belirgin düşüşü sergiledi.



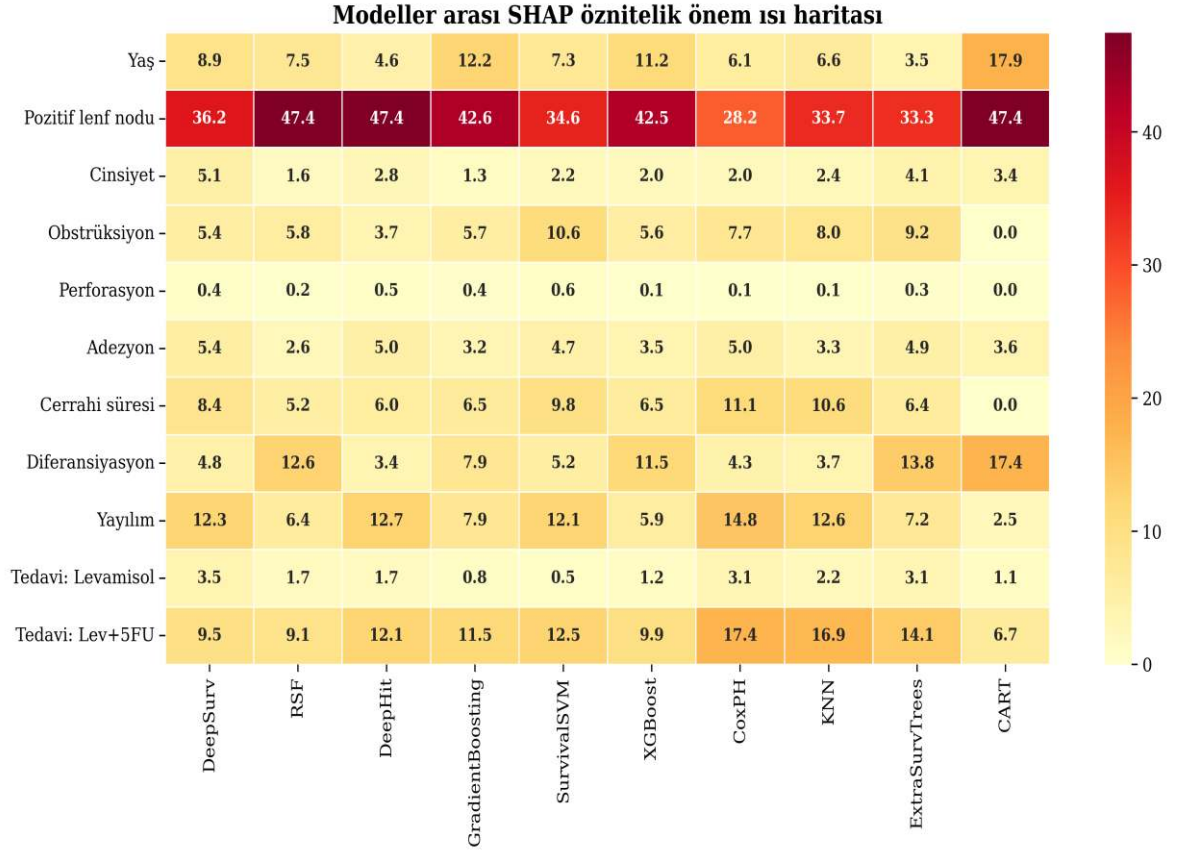
Şekil 4.4. Tüm modellere ait kümülatif/dinamik AUC grafiği

Normalize SHAP değerleri Tablo 4.3'te verildi. Tablo 4.3'e göre pozitif lenf düğümü sayısı tüm modellerde açık ara en yüksek öneme sahiptir (%33-47). Kalın bağırsak perforasyonu tüm modellerde en düşük değerdedir. Lev+5FU tedavi grubu (Gözlem referans) CoxPH'de %17.4, KNN'de %16.9 iken ağaç tabanlı modellerde daha düşük kaldı.

Tablo 4.3. Normalize SHAP değerleri

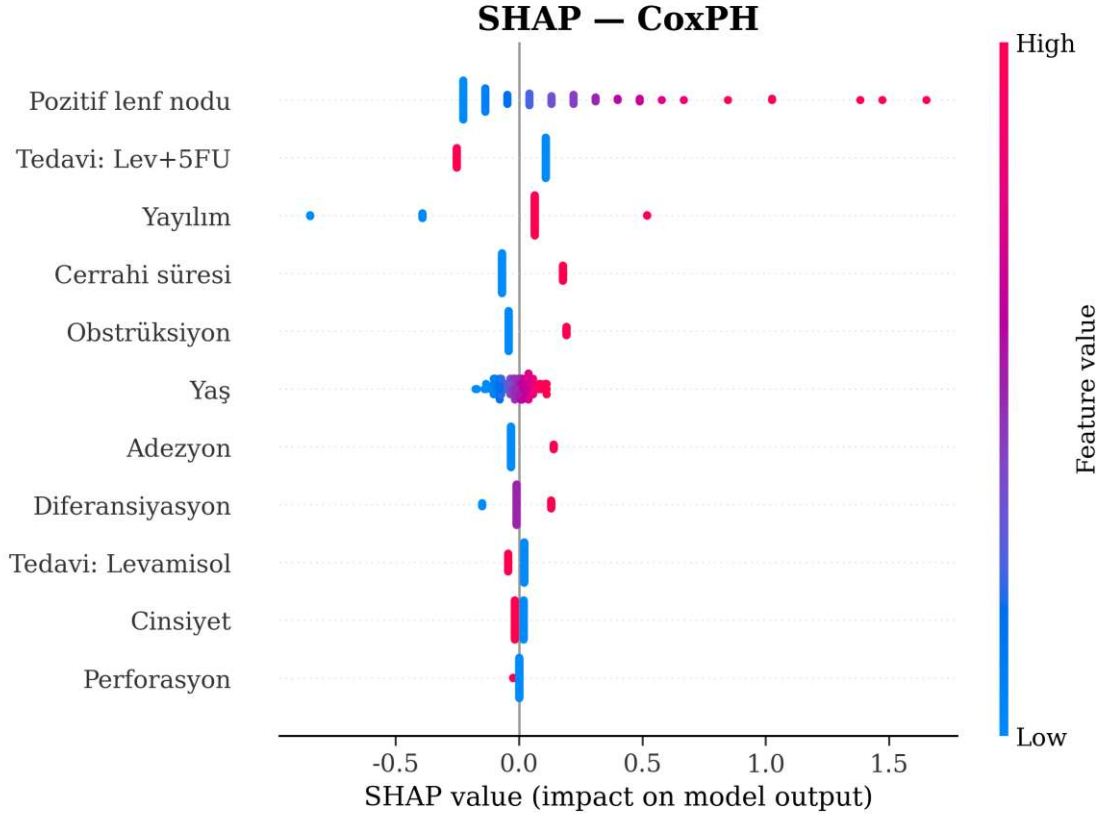
Öznitelik	DeepSurv	SurvivalSVM	RSF	GradientBoosting	ExtraSurvTrees	CoxPH	KNN	XGBoost	DeepHit	CART
Pozitif lenf nodu	36.2	34.6	47.4	42.6	33.3	28.2	33.7	42.5	47.4	47.5
Yayılım	12.3	12.1	6.4	7.9	7.2	14.8	12.6	5.9	12.7	2.5
Tedavi: Lev+5FU	9.5	12.5	9.1	11.5	14.1	17.4	16.9	9.9	12.1	6.7
Yaş	8.9	7.3	7.5	12.2	3.5	6.1	6.6	11.2	4.6	17.9
Cerrahi süresi	8.4	9.8	5.2	6.5	6.4	11.1	10.6	6.5	6.0	0.0
Obstrüksiyon	5.4	10.6	5.8	5.7	9.2	7.7	8.0	5.6	3.7	0.0
Adezyon	5.4	4.7	2.6	3.2	4.9	5.0	3.3	3.5	5.0	3.6
Cinsiyet	5.1	2.2	1.6	1.3	4.1	2.0	2.4	2.0	2.8	3.4
Diferansiyasyon	4.8	5.2	12.6	7.9	13.8	4.3	3.7	11.5	3.4	17.4
Tedavi: Levamisol	3.5	0.6	1.7	0.8	3.1	3.1	2.2	1.2	1.7	1.1
Perforasyon	0.4	0.6	0.1	0.4	0.3	0.1	0.1	0.1	0.5	0.0

Şekil 4.5'te pozitif lenf düğümü sayısı satırı tüm modellerde en koyu renk ile gösterildi. Kalın bağırsak perforasyonu satırı tüm modellerde açık renktir.



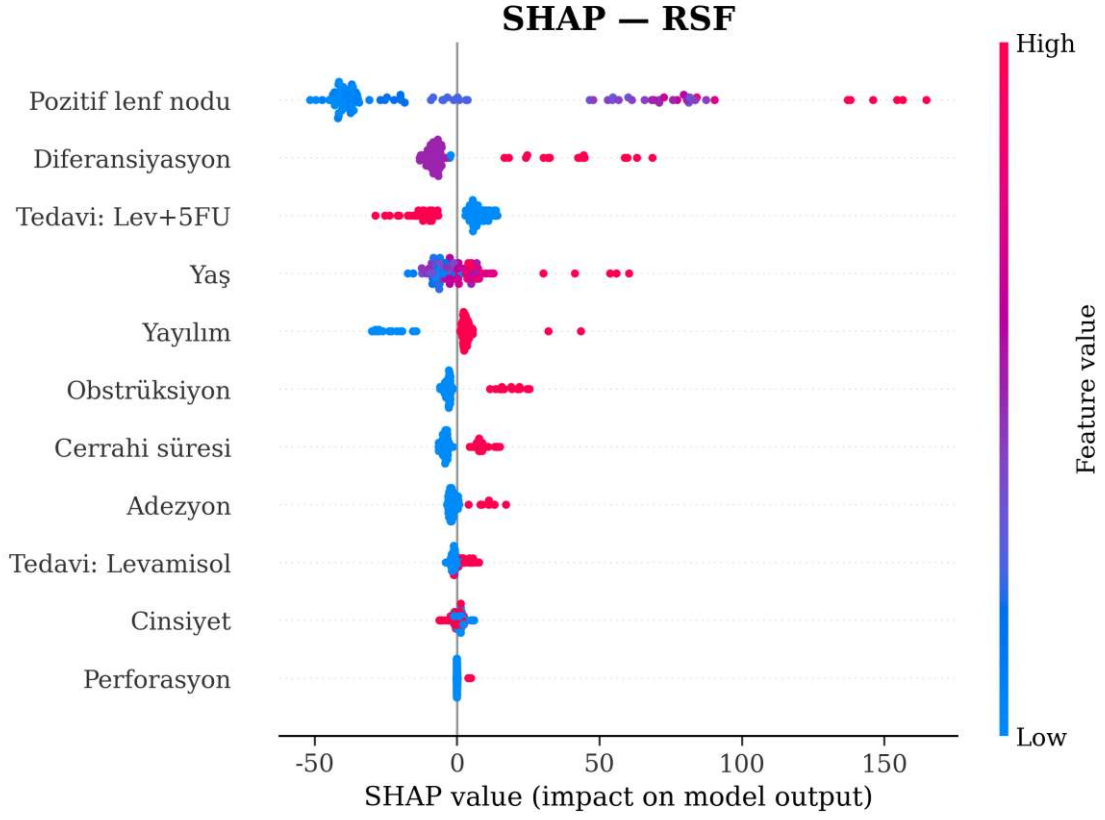
Şekil 4.5. Modeller arası SHAP öznitelik önem ısı haritası

Şekil 4.6'da CoxPH modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%28.2), Lev+5FU tedavisi (%17.4) ve yayılım düzeyi (%14.8) olarak belirlendi.



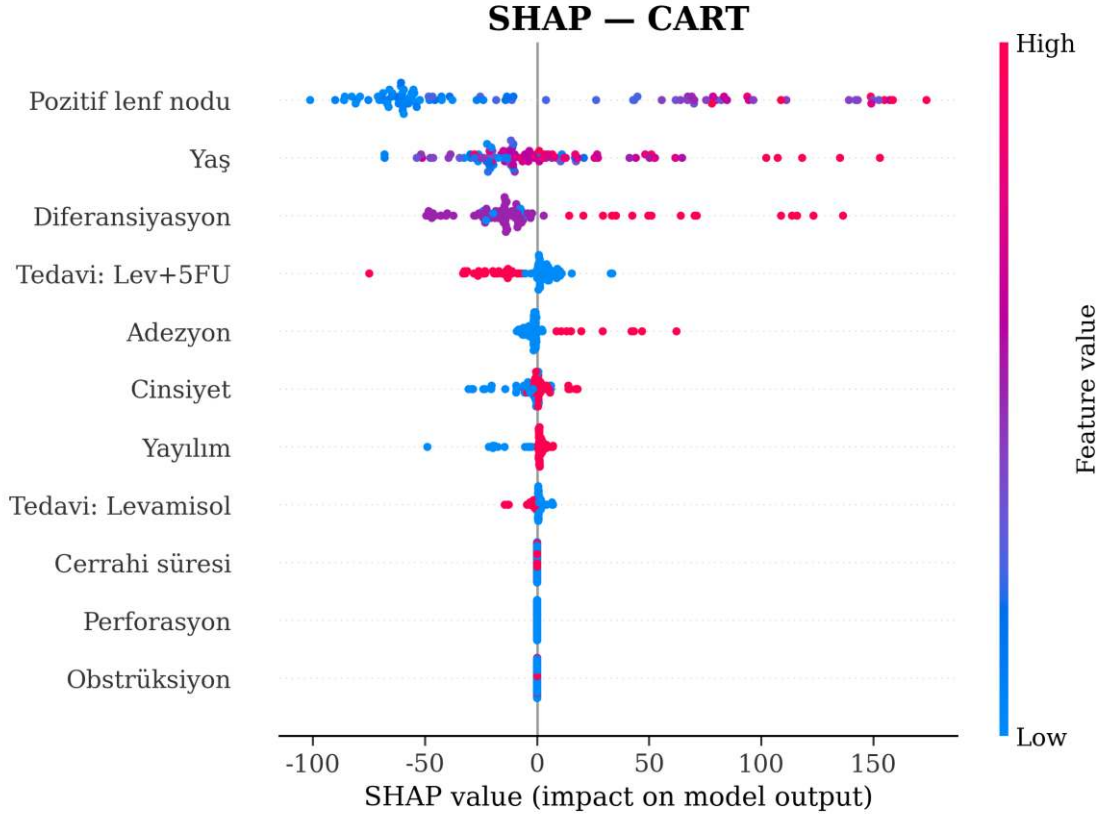
Şekil 4.6. CoxPH modeline ait SHAP özet grafiği

Şekil 4.7'de RSF modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%47.4), diferansiyasyon derecesi (%12.6) ve Lev+5FU tedavisi (%9.1) olarak belirlendi.



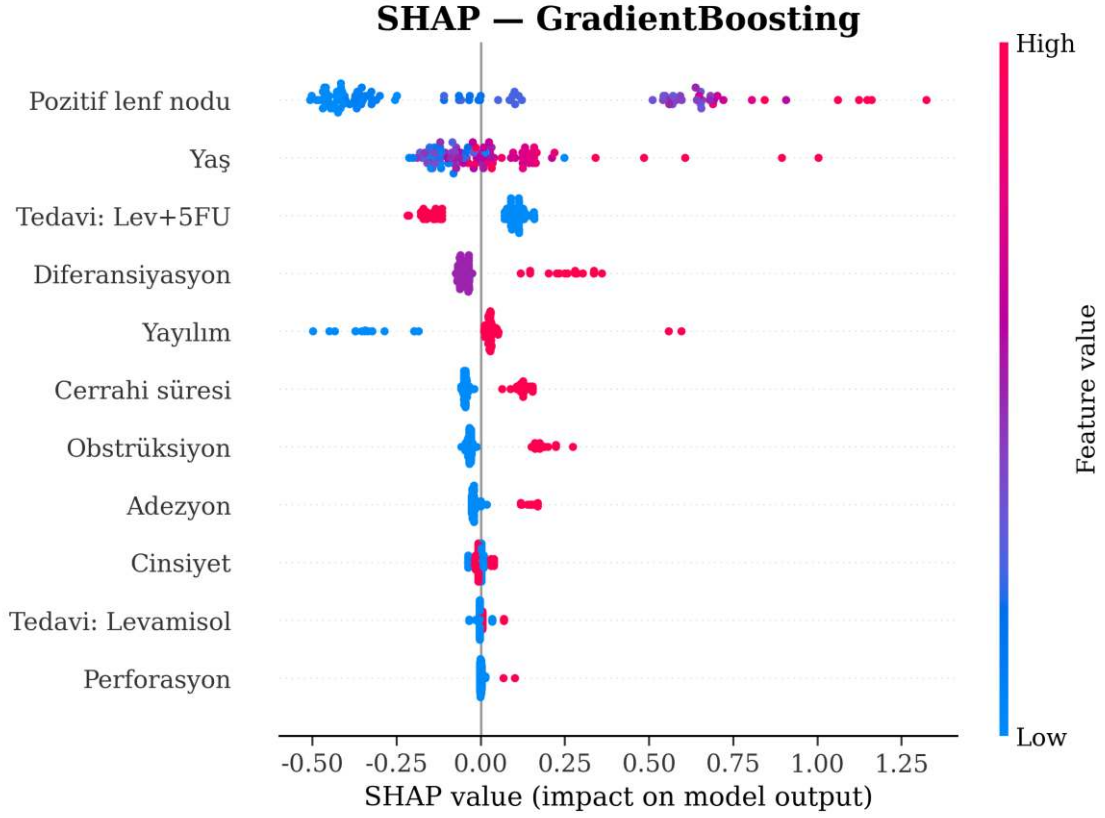
Şekil 4.7. RSF modeline ait SHAP özet grafiği

Şekil 4.8'de CART modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%47.5), yaş (%17.9) ve diferansiyasyon derecesi (%17.4) olarak belirlendi.



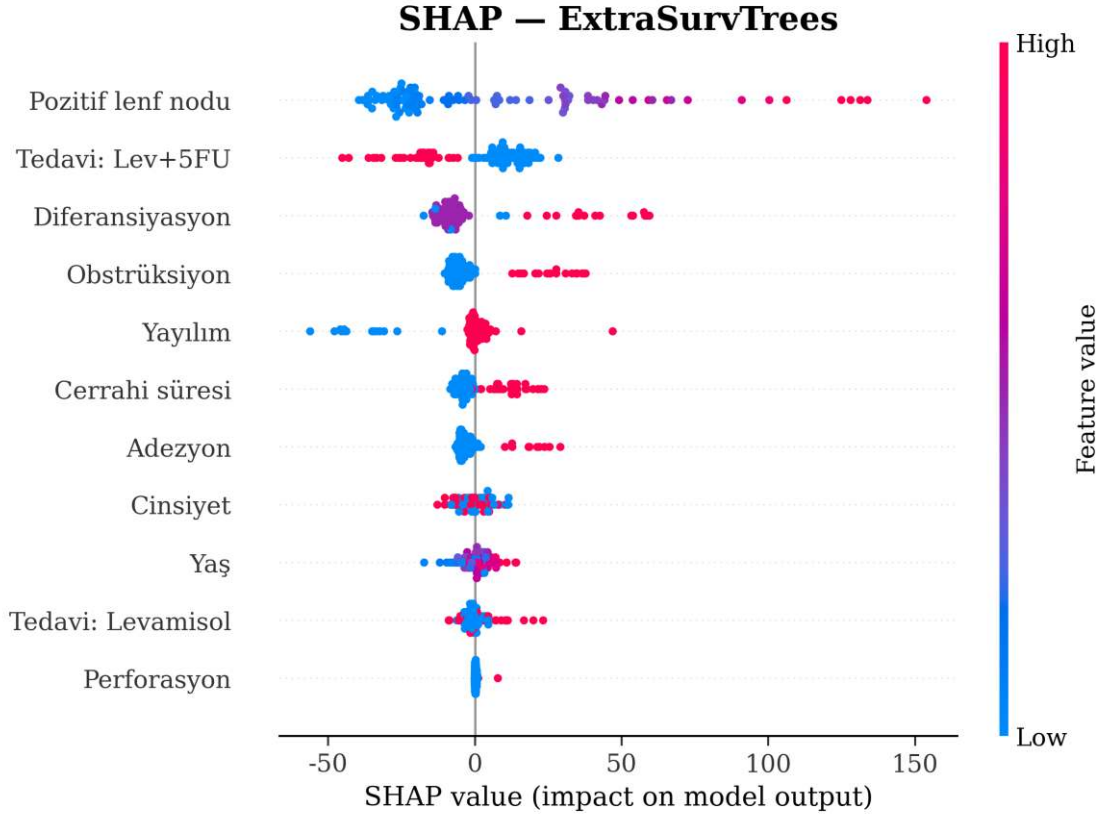
Şekil 4.8. CART modeline ait SHAP özet grafiği

Şekil 4.9'da GradientBoosting modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%42.6), yaş (%12.2) ve Lev+5FU tedavisi (%11.5) olarak belirlendi.



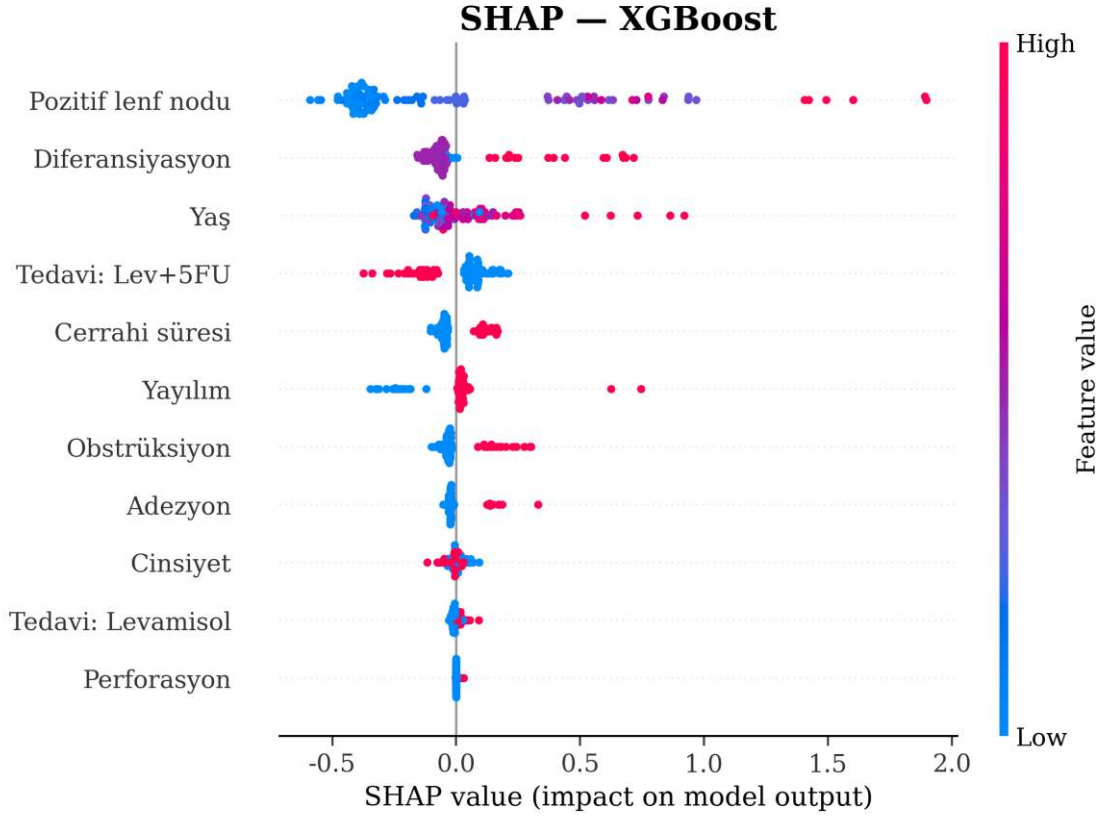
Şekil 4.9. GradientBoosting modeline ait SHAP özet grafiği

Şekil 4.10'da ExtraSurvTrees modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%33.3), Lev+5FU tedavisi (%14.1) ve diferansiyasyon derecesi (%13.8) olarak belirlendi.



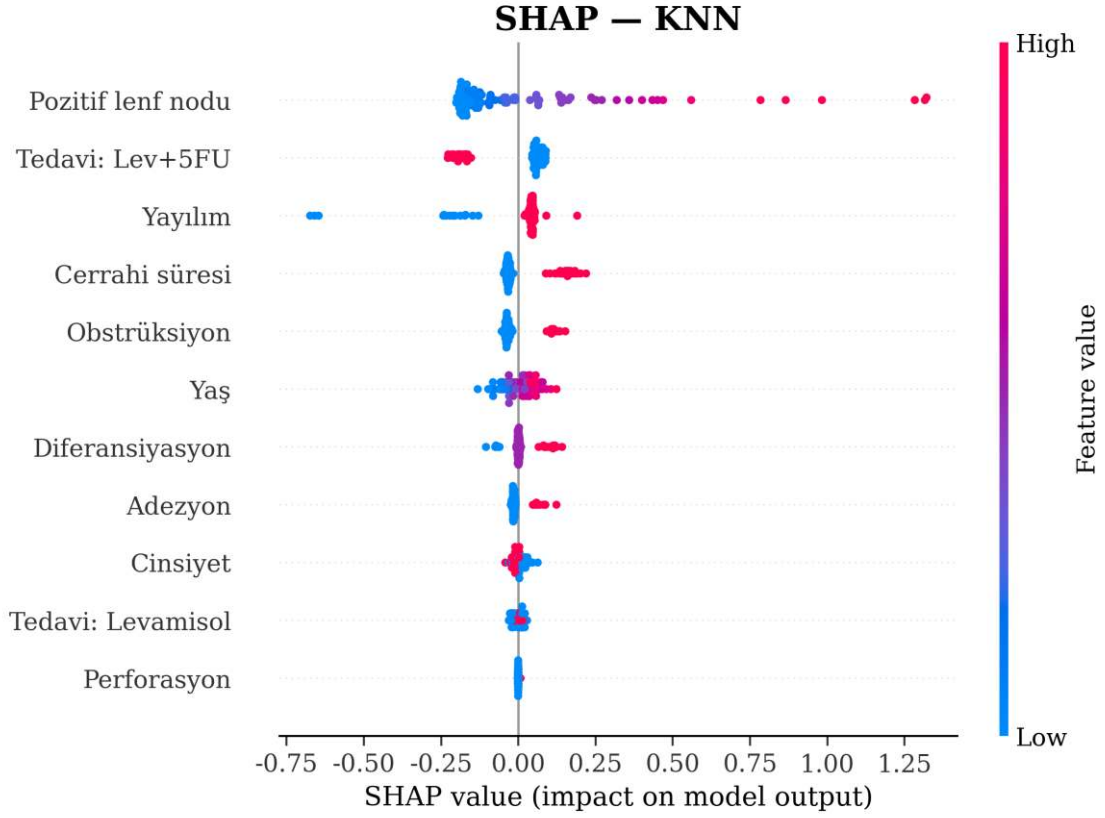
Şekil 4.10. ExtraSurvTrees modeline ait SHAP özet grafiği

Şekil 4.11'de XGBoost modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%42.5), diferansiyasyon derecesi (%11.5) ve yaş (%11.2) olarak belirlendi.



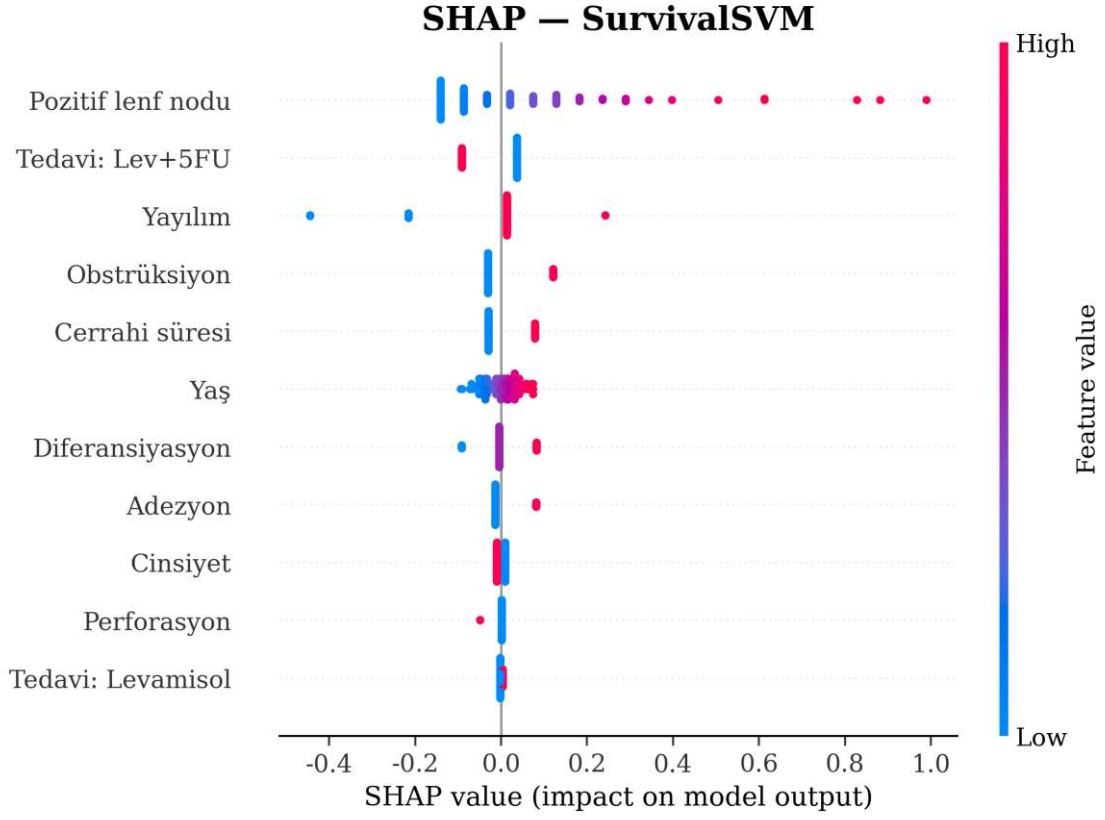
Şekil 4.11. XGBoost modeline ait SHAP özet grafiği

Şekil 4.12'de KNN modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%33.7), Lev+5FU tedavisi (%16.9) ve yayılım düzeyi (%12.6) olarak belirlendi.



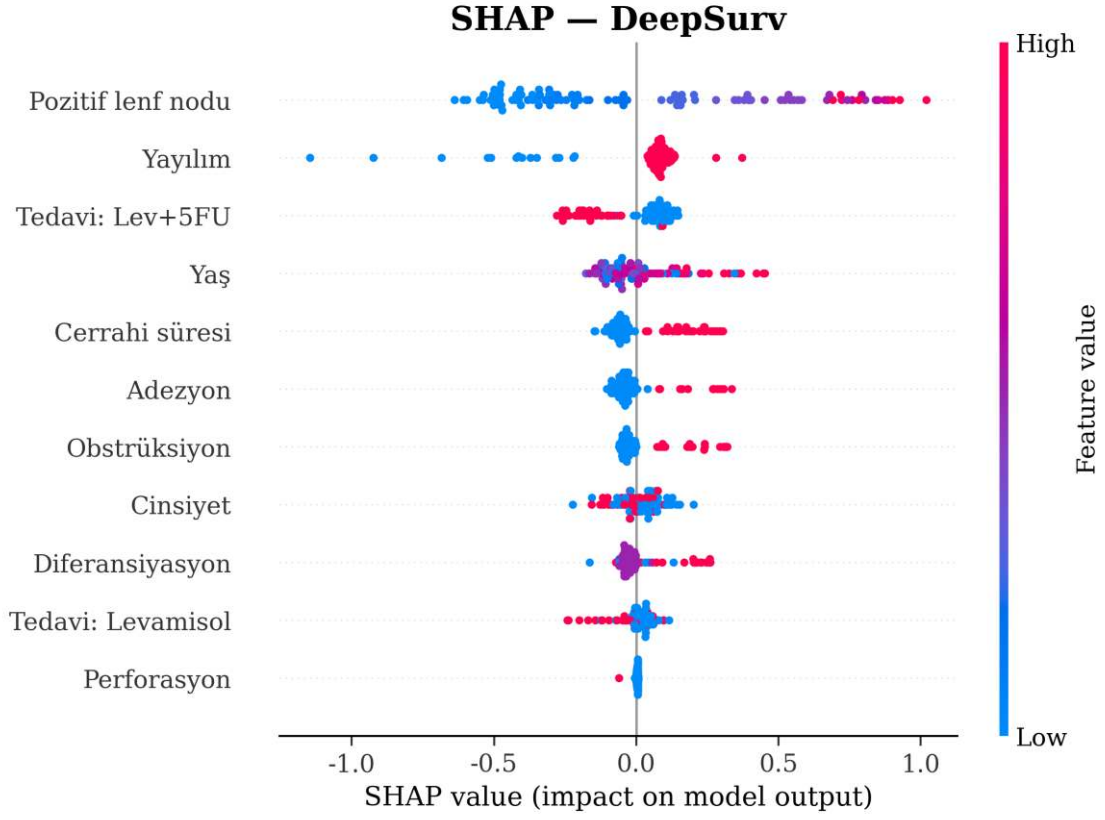
Şekil 4.12. KNN modeline ait SHAP özet grafiği

Şekil 4.13'te SurvivalSVM modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%34.6), Lev+5FU tedavisi (%12.5) ve yayılım düzeyi (%12.1) olarak belirlendi.



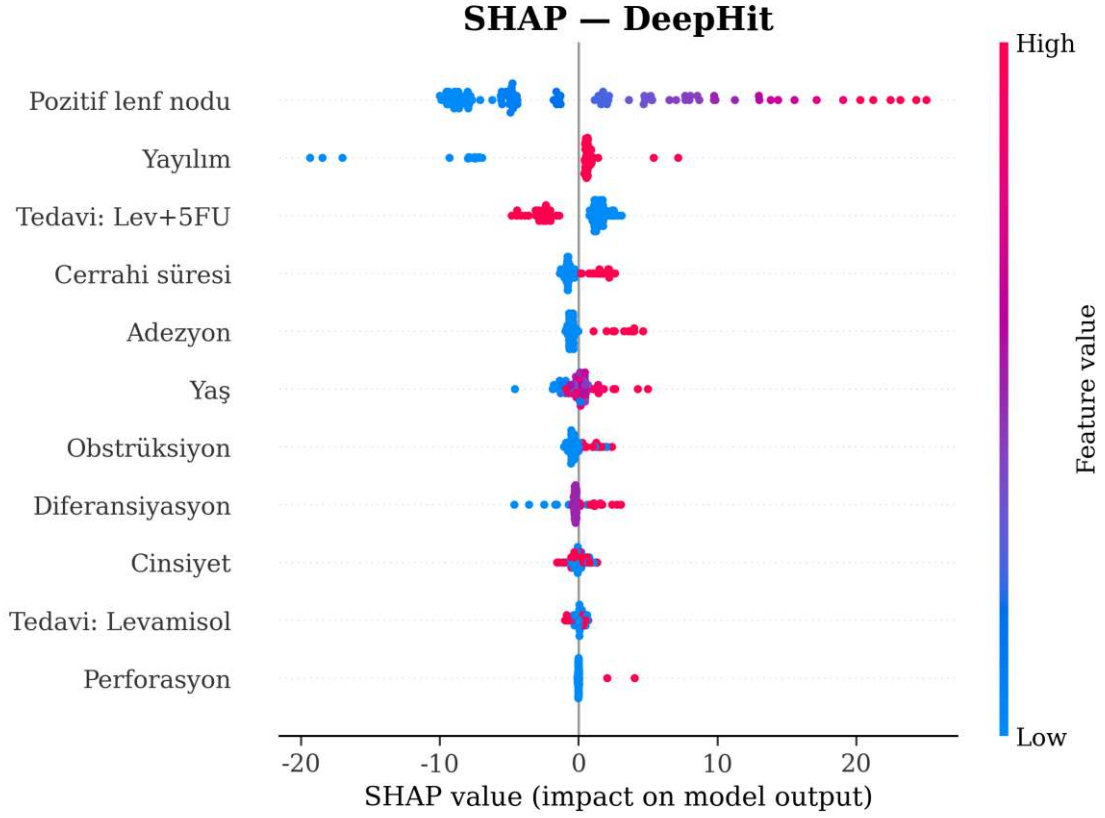
Şekil 4.13. SurvivalSVM modeline ait SHAP özet grafiği

Şekil 4.14'te DeepSurv modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%36.2), yayılım düzeyi (%12.3) ve Lev+5FU tedavisi (%9.5) olarak belirlendi.



Şekil 4.14. DeepSurv modeline ait SHAP özet grafiği

Şekil 4.15'te DeepHit modeli için en yüksek öneme sahip üç değişken sırasıyla pozitif lenf düğümü sayısı (%47.4), yayılım düzeyi (%12.7) ve Lev+5FU tedavisi (%12.1) olarak belirlendi.

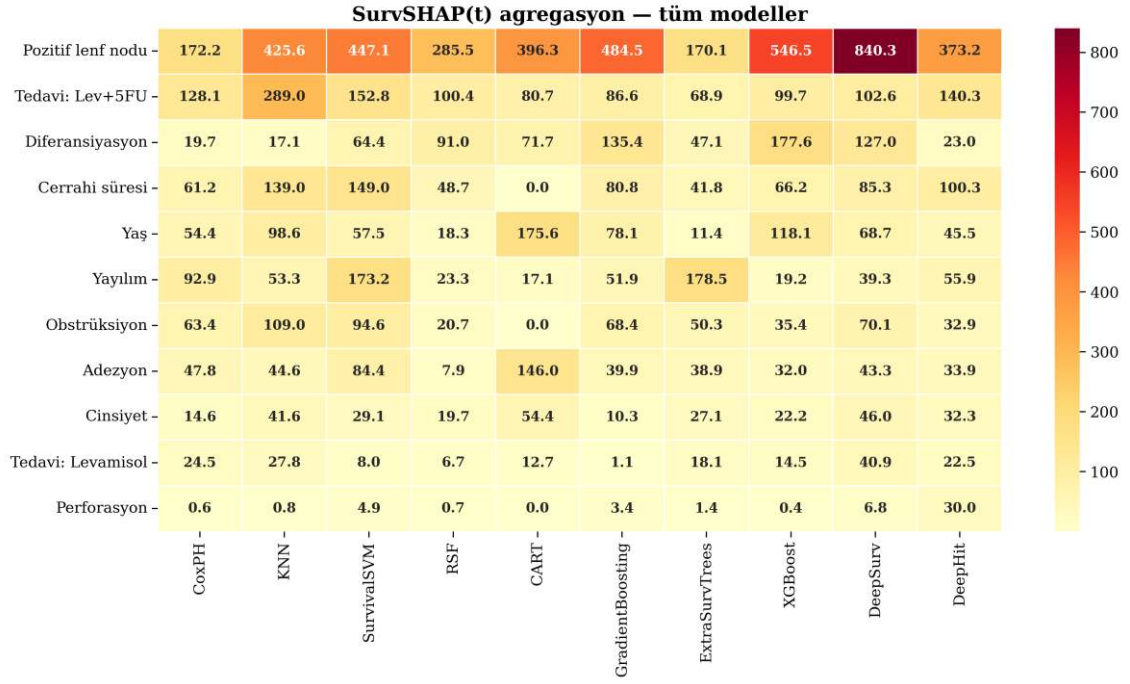


Şekil 4.15. DeepHit modeline ait SHAP özet grafiği

SurvSHAP(t) agregasyon değerleri Tablo 4.4'te verildi. Tablo 4.4'e göre pozitif lenf düğümü sayısı tüm modellerde en yüksek agregasyon değerine sahiptir; ikinci sırada Lev+5FU tedavi grubu yer aldı. Bu sıralama Tablo 4.3'teki SHAP sonuçlarıyla tutarlıdır.

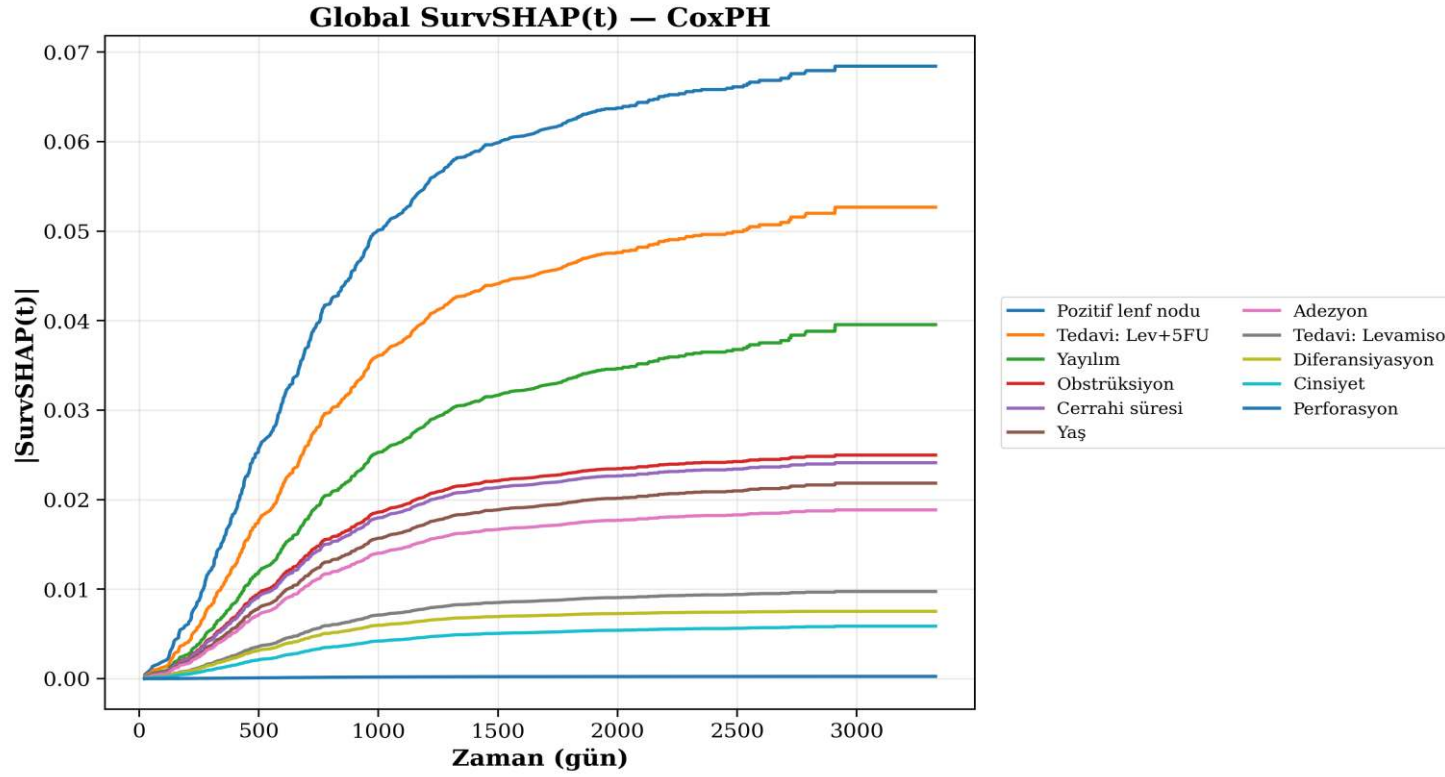
Tablo 4.4. SurvSHAP(t) agregasyonu yapılmış değerler

Öznitelik	CoxP H	KNN	SurvivalS VM	RSF	CART	GradientBoo sting	ExtraSurv Trees	XGBo ost	DeepS urv	Deep Hit
Pozitif lenf nodu	172.20 27	425.55 61	447.0681	285.50 28	396.27 21	484.5094	170.0922	546.47 10	840.27 62	373.19 96
Tedavi: Lev+5FU	128.06 07	288.99 97	152.7850	100.36 34	80.700 5	86.6452	68.9248	99.684 1	102.58 71	140.33 04
Yayılım	92.923 8	53.300 5	173.2058	23.294 4	17.112 2	51.9375	178.4962	19.200 1	39.291 5	55.946 4
Obstrüksiy on	63.380 4	108.99 51	94.5743	20.697 0	0.0001	68.3716	50.2717	35.369 6	70.107 3	32.919 3
Cerrahi süresi	61.198 4	139.03 64	149.0212	48.726 2	0.0001	80.8076	41.7927	66.222 8	85.255 7	100.31 85
Yaş	54.385 6	98.629 0	57.4758	18.276 8	175.58 39	78.0777	11.3888	118.14 78	68.699 3	45.491 9
Adezyon	47.793 8	44.646 8	84.4357	7.9152	145.98 32	39.9498	38.8782	32.010 1	43.269 2	33.908 6
Tedavi: Levamisol	24.459 6	27.810 9	8.0026	6.6540	12.670 3	1.1312	18.0551	14.540 7	40.911 7	22.548 3
Diferansiya syon	19.735 8	17.136 4	64.3835	90.980 1	71.656 5	135.4313	47.0572	177.61 58	126.99 29	23.043 8
Cinsiyet	14.584 7	41.643 0	29.1079	19.726 9	54.386 4	10.2858	27.0652	22.210 2	46.008 0	32.255 8
Perforasyo n	0.6474	0.7518	4.9100	0.6609	0.0000	3.4352	1.4339	0.4335	6.7880	30.001 7



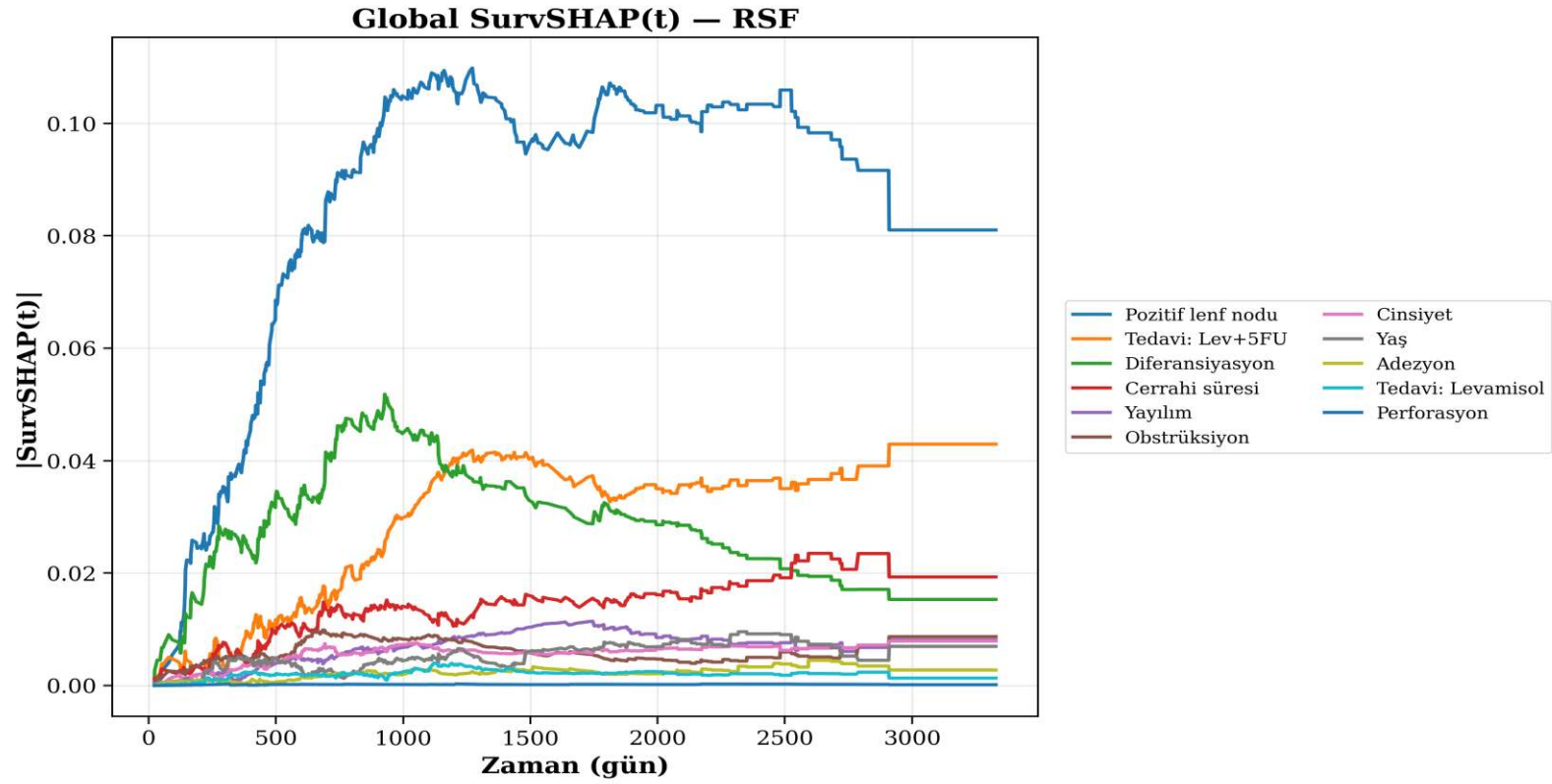
Şekil 4.16. Modeller arası SurvSHAP(t) ısı haritası

Şekil 4.17'de CoxPH modeli için zamana bağlı öznelilik etkileri gösterildi. Pozitif lenf düğümü sayısı tüm zaman noktalarında en yüksek mutlak SurvSHAP(t) değerine sahiptir. Değişkenlerin göreceli sıralaması zaman boyunca sabit kaldı.



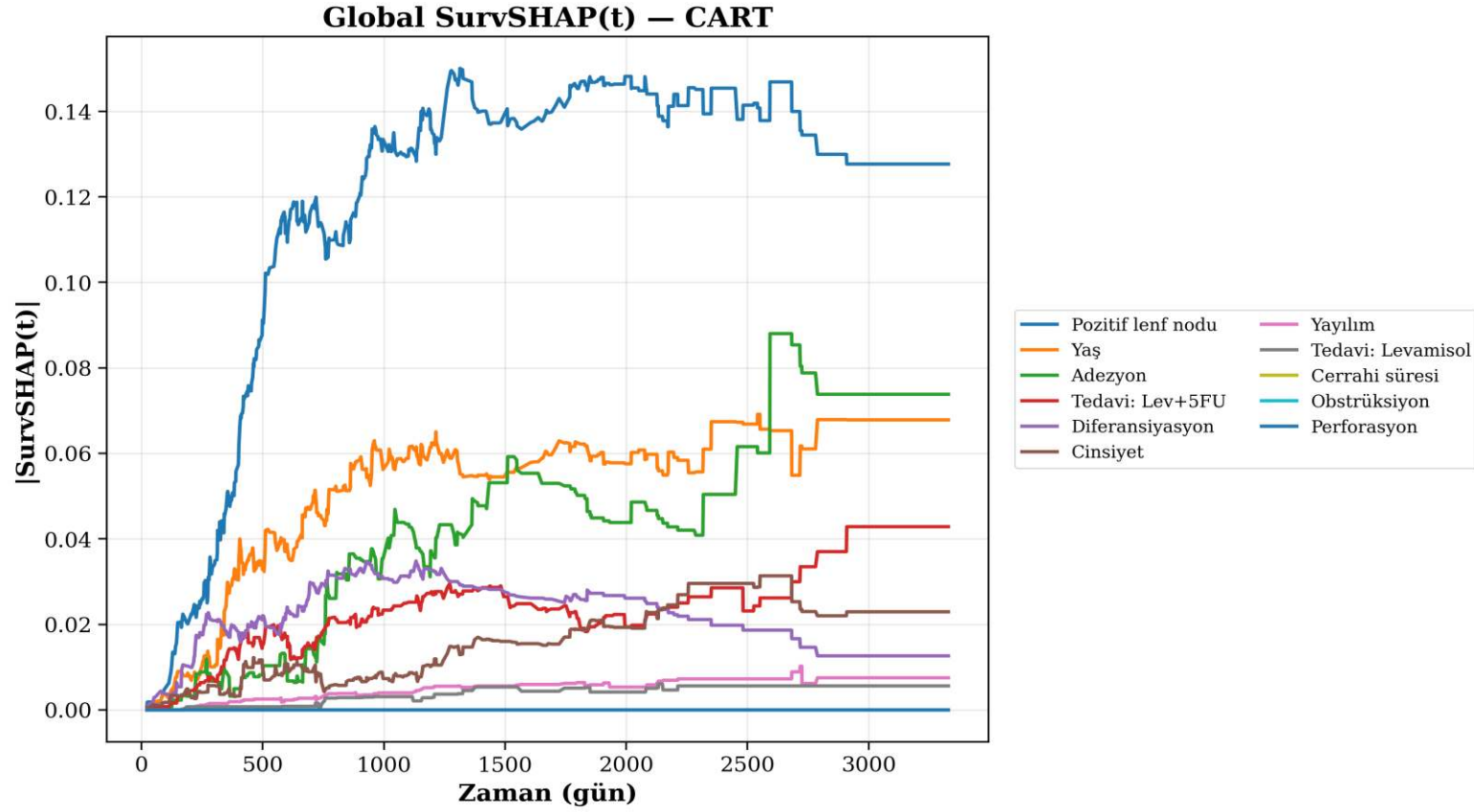
Şekil 4.17. CoxPH modeline ait global SurvSHAP(t) grafiği

Şekil 4.18'de RSF modeli için zamana bağlı öznitelik etkileri gösterildi. Pozitif lenf düğümü sayısı ilk 1000 günde hızla artmakta, sonrasında platoya ulaştı. Lev+5FU tedavisi ikinci sırada yer aldı.



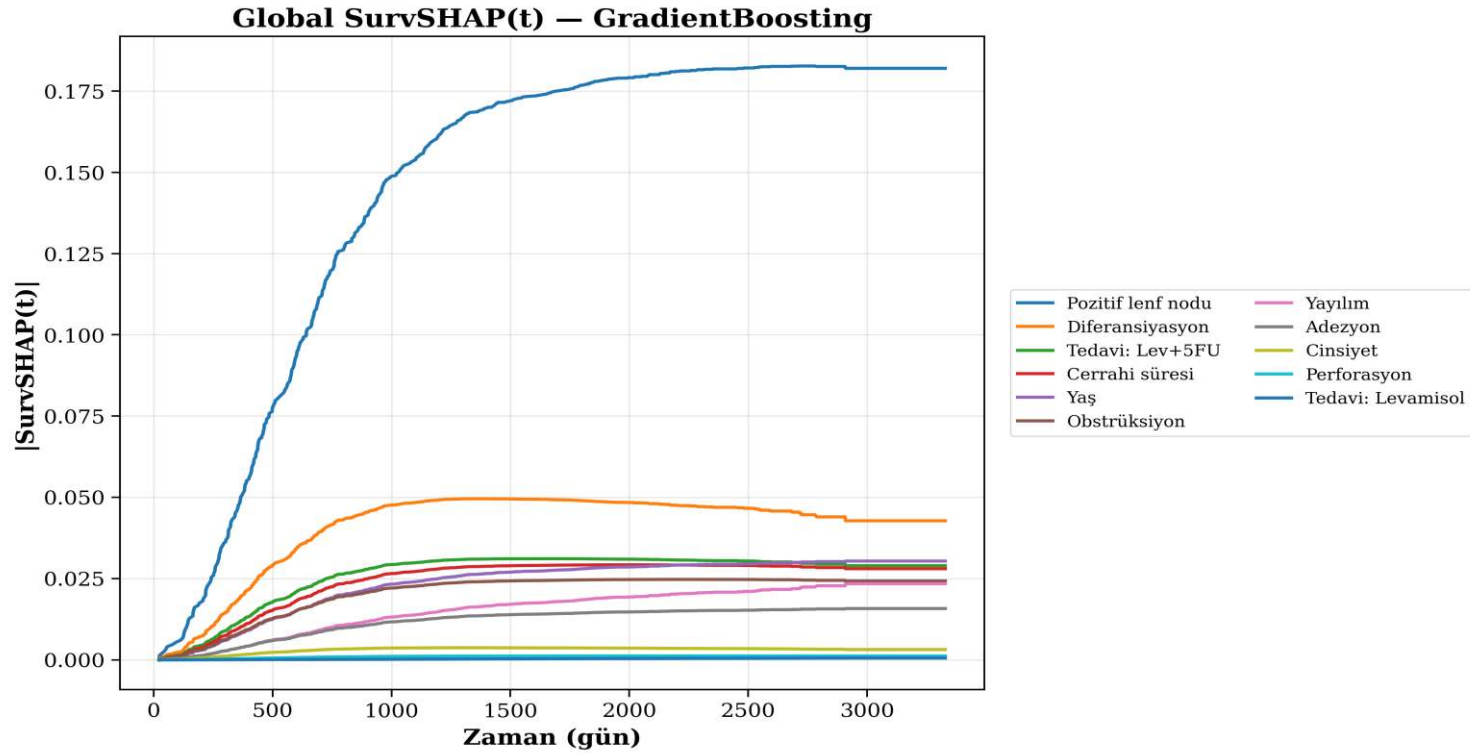
Şekil 4.18. RSF modeline ait global SurvSHAP(t) grafiği

Şekil 4.19'da CART modeli için zamana bağlı öznitelik etkileri gösterildi. Pozitif lenf düğümü sayısı ve yaş basamak şeklinde etki gösterdi. Değişken sayısı az olduğundan eğriler ayrık bir görünüm sergiledi.



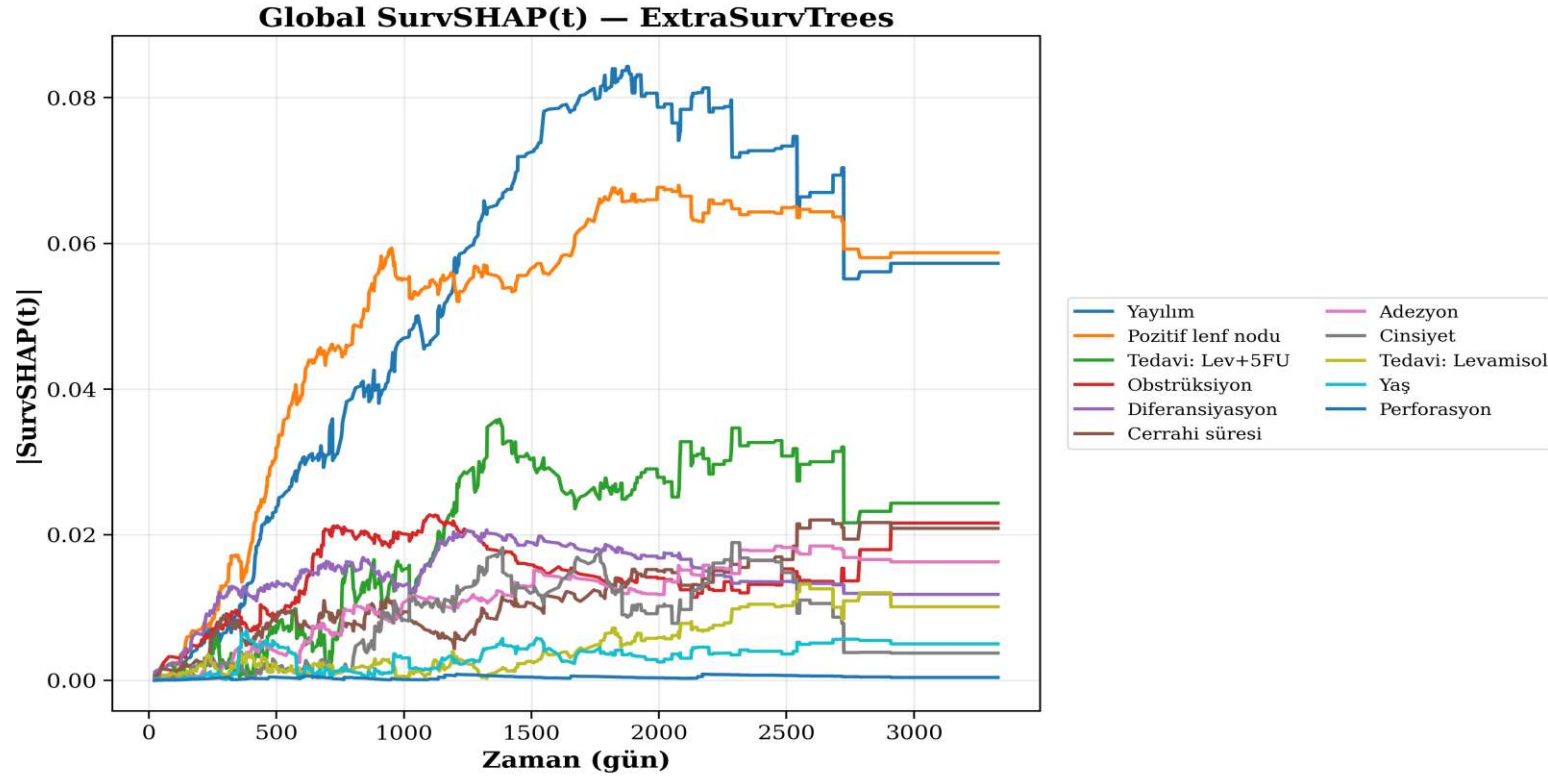
Şekil 4.19. CART modeline ait global SurvSHAP(t) grafiği

Şekil 4.20'de GradientBoosting modeli için zamana bağlı öznelilik etkileri gösterildi. Pozitif lenf düğümü sayısı zamana bağlı olarak sürekli artan bir etki gösterdi. Yayılım düzeyi erken dönemde daha belirgindir.



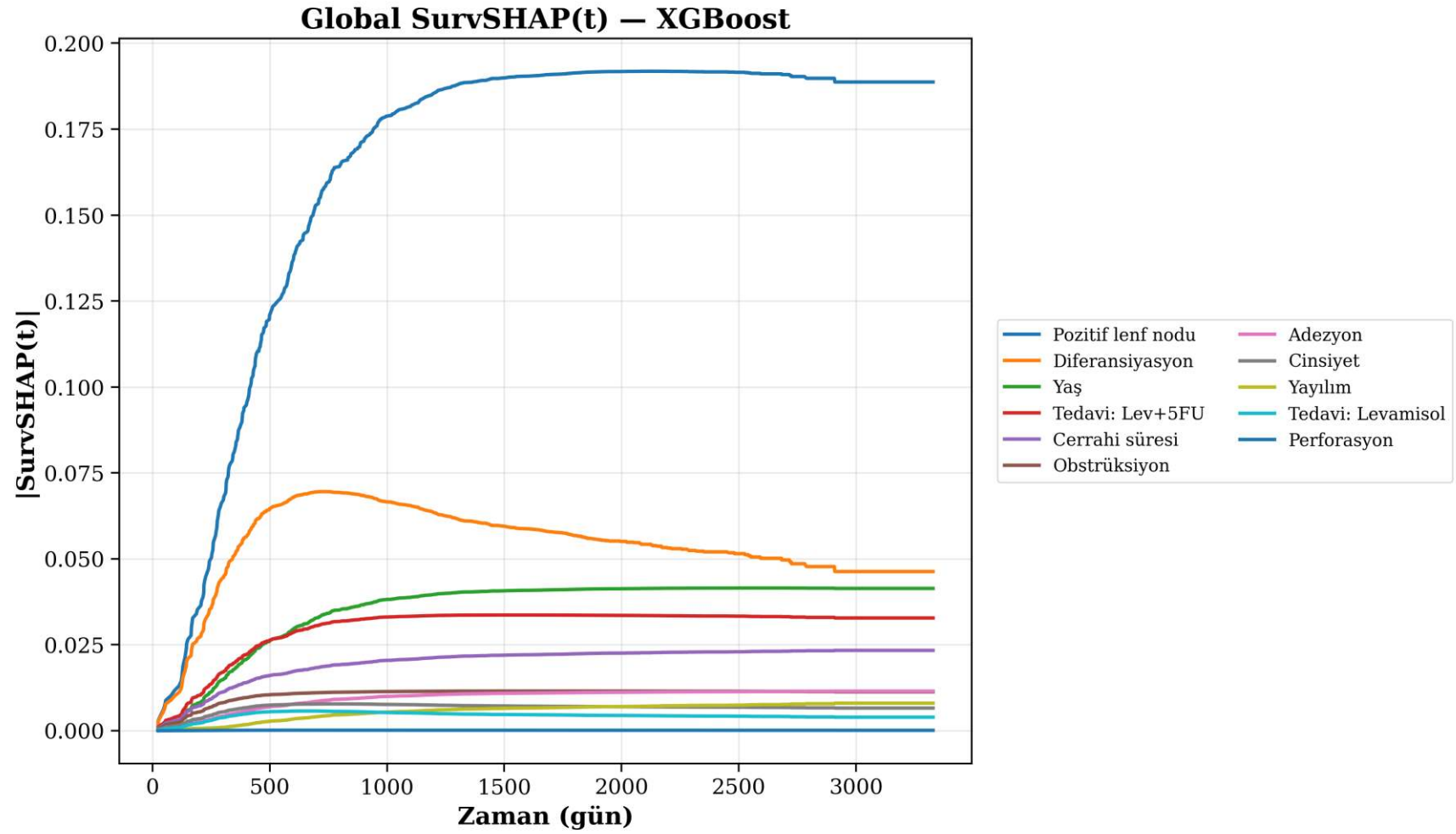
Şekil 4.20. GradientBoosting modeline ait global SurvSHAP(t) grafiği

Şekil 4.21'de ExtraSurvTrees modeli için zamana bağlı öznelilik etkileri gösterildi. Yayılım düzeyi ve pozitif lenf düğümü sayısı tüm zaman noktalarında baskın konumdadır. Komşu organlara yapışma beklenenden yüksek etki gösterdi.



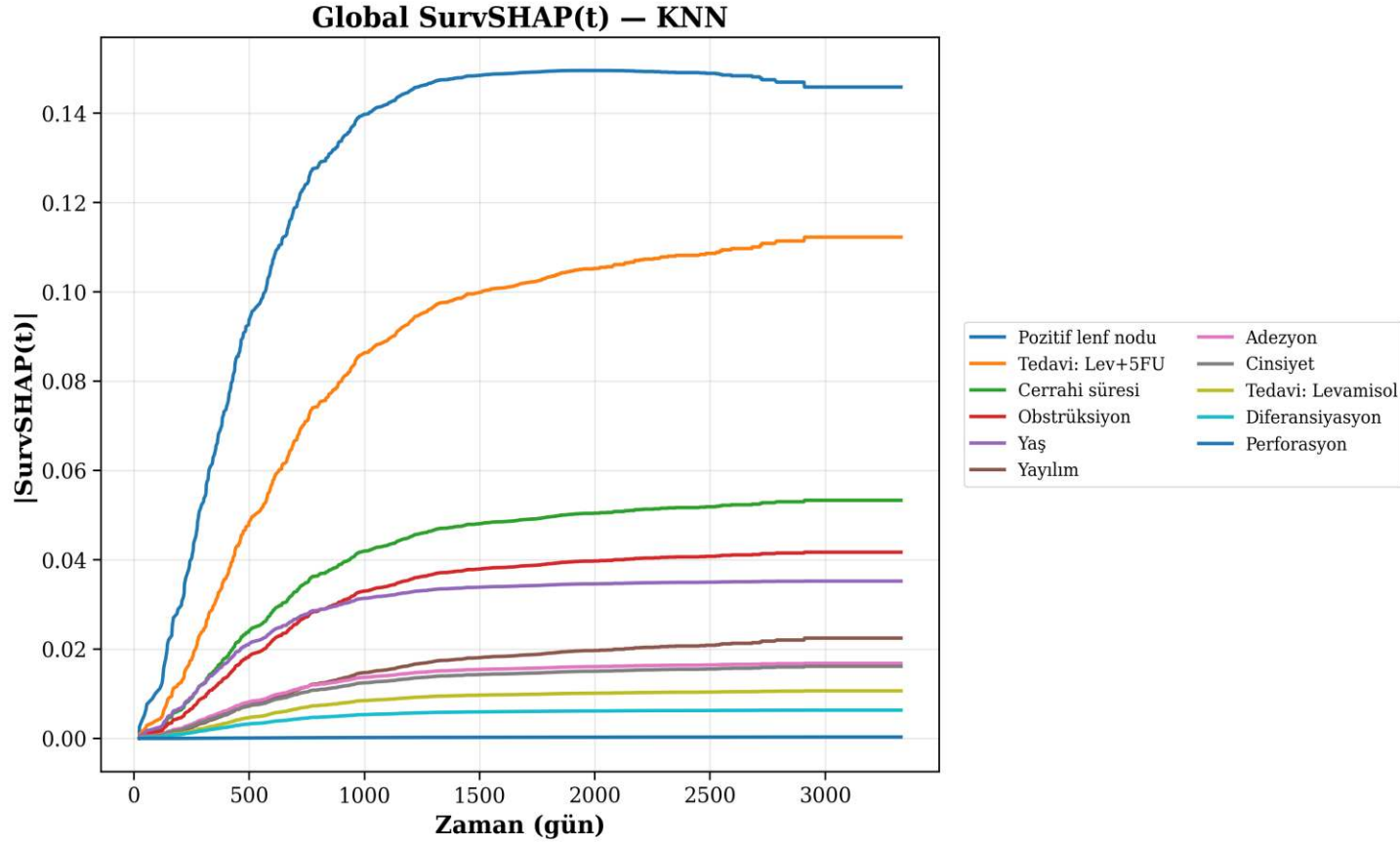
Şekil 4.21. ExtraSurvTrees modeline ait global SurvSHAP(t) grafiği

Şekil 4.22'de XGBoost modeli için zamana bağlı öznelik etkileri gösterildi. Pozitif lenf düğümü sayısı ilk 500 günde dik yükseliş göstermekte ve tüm zaman dilimlerinde baskın değişken olarak kaldı.



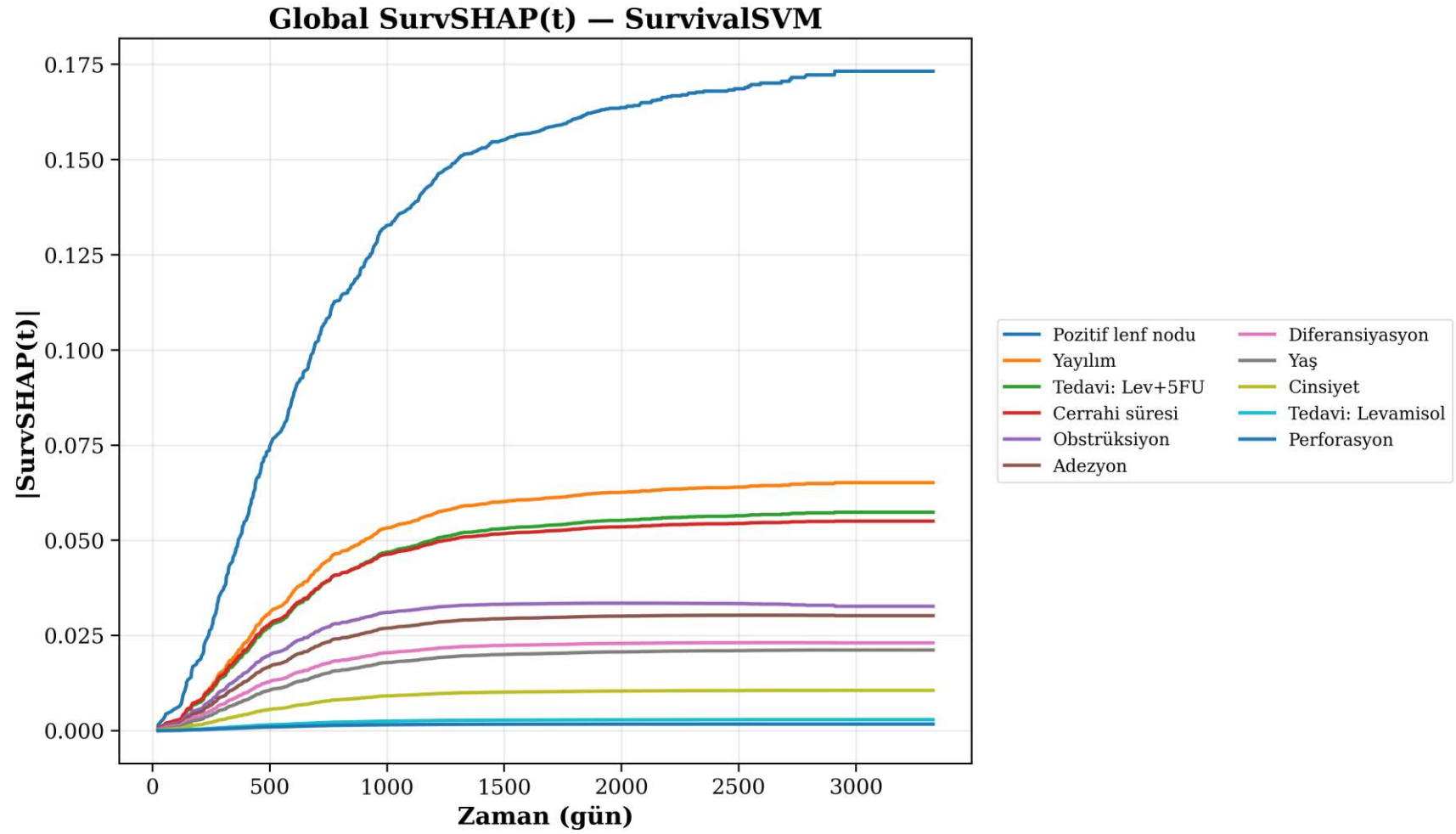
Şekil 4.22. XGBoost modeline ait global SurvSHAP(t) grafiği

Şekil 4.23'te KNN modeli için zamana bağlı öznelilik etkileri gösterildi. Tüm değişkenlerin etkisi zamana bağlı olarak monoton artış gösterdi. Pozitif lenf düğümü sayısı en yüksek değere sahiptir.



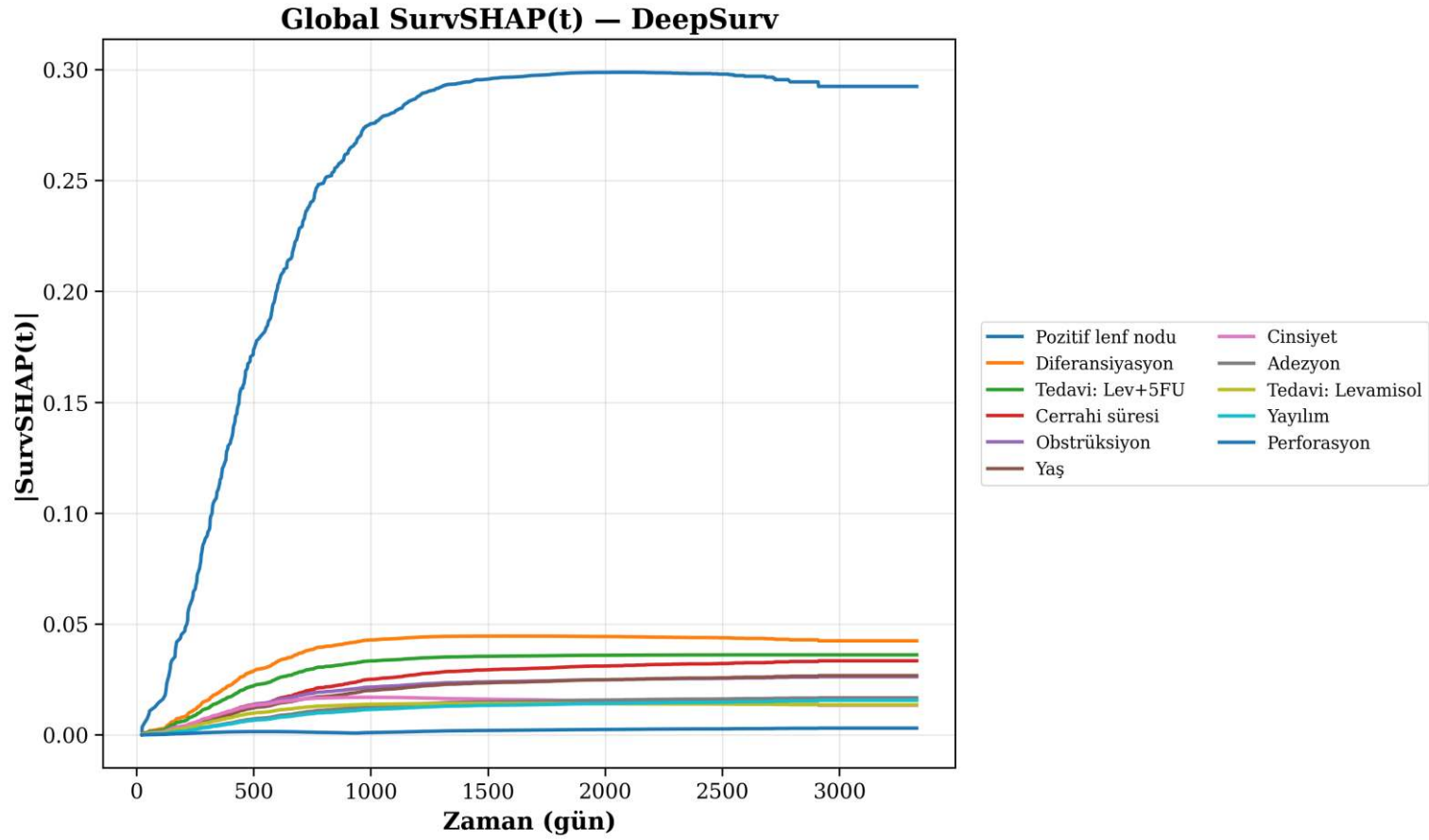
Şekil 4.23. KNN modeline ait global SurvSHAP(t) grafiği

Şekil 4.24'te SurvivalSVM modeli için zamana bağlı öznelilik etkileri gösterildi. Pozitif lenf düğümü sayısı ve yayılım düzeyi erken dönemde belirgin, uzun vadede azalan etki gösterdi.



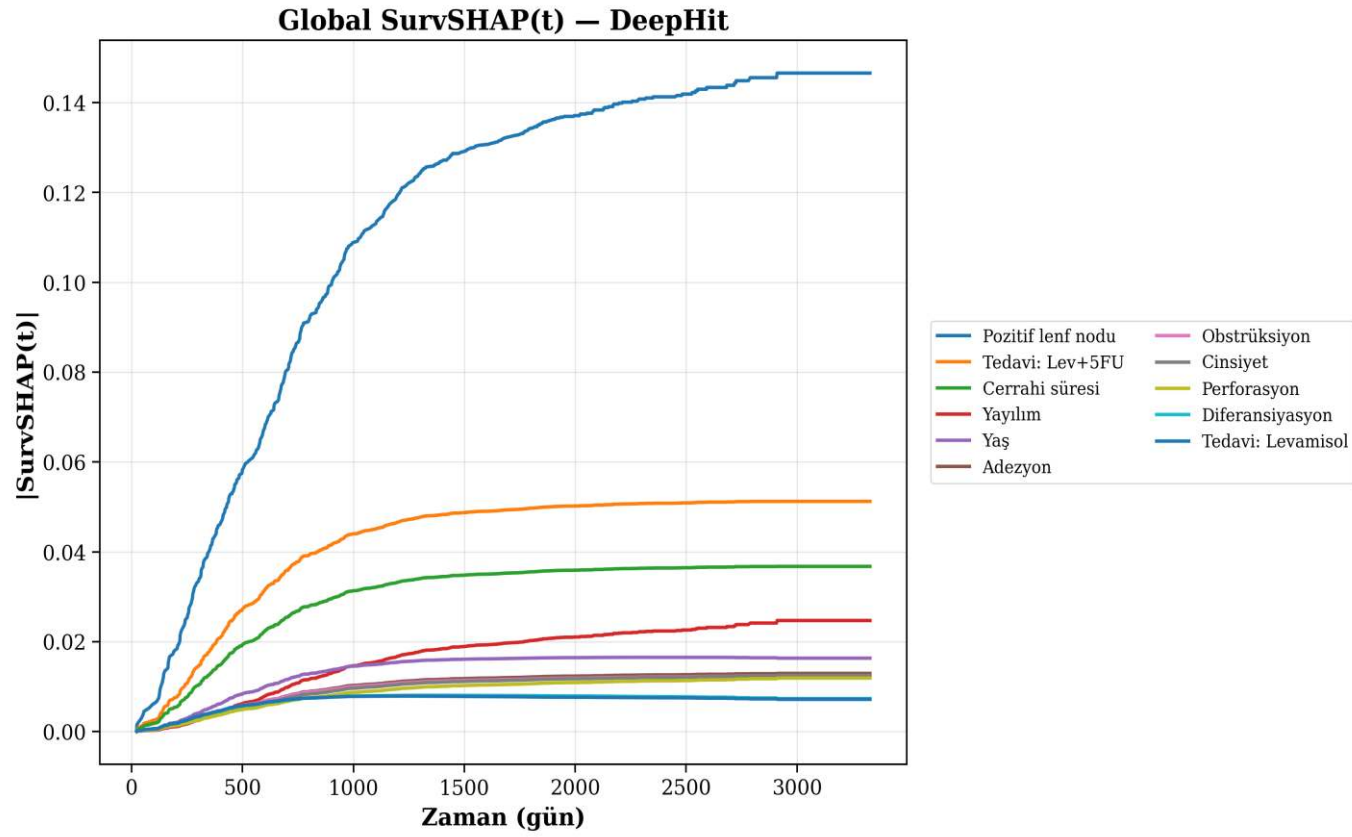
Şekil 4.24. SurvivalSVM modeline ait global SurvSHAP(t) grafiği

Şekil 4.25'te DeepSurv modeli için zamana bağlı öznel etkileri gösterildi. Pozitif lenf düğümü sayısı diğer tüm değişkenlerden belirgin biçimde ayrılarak en yüksek etkiyi göstermekte; diferansiyasyon derecesi ve Lev+5FU tedavisi ikincil konumda yer aldı.



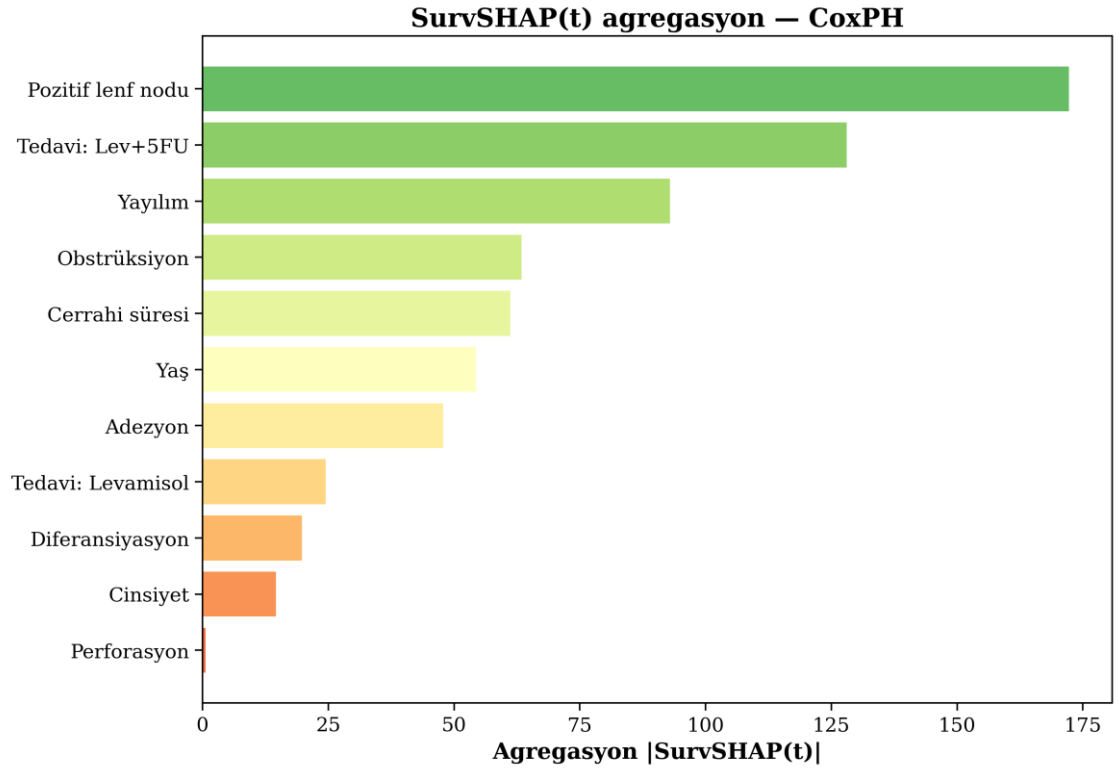
Şekil 4.25. DeepSurv modeline ait global SurvSHAP(t) grafiği

Şekil 4.26'da DeepHit modeli için zamana bağlı öznitelik etkileri gösterildi. Pozitif lenf düğümü sayısı diğer tüm değişkenlerden belirgin şekilde ayrıştı. Eğriler basamaklı bir görünüm sergiledi.



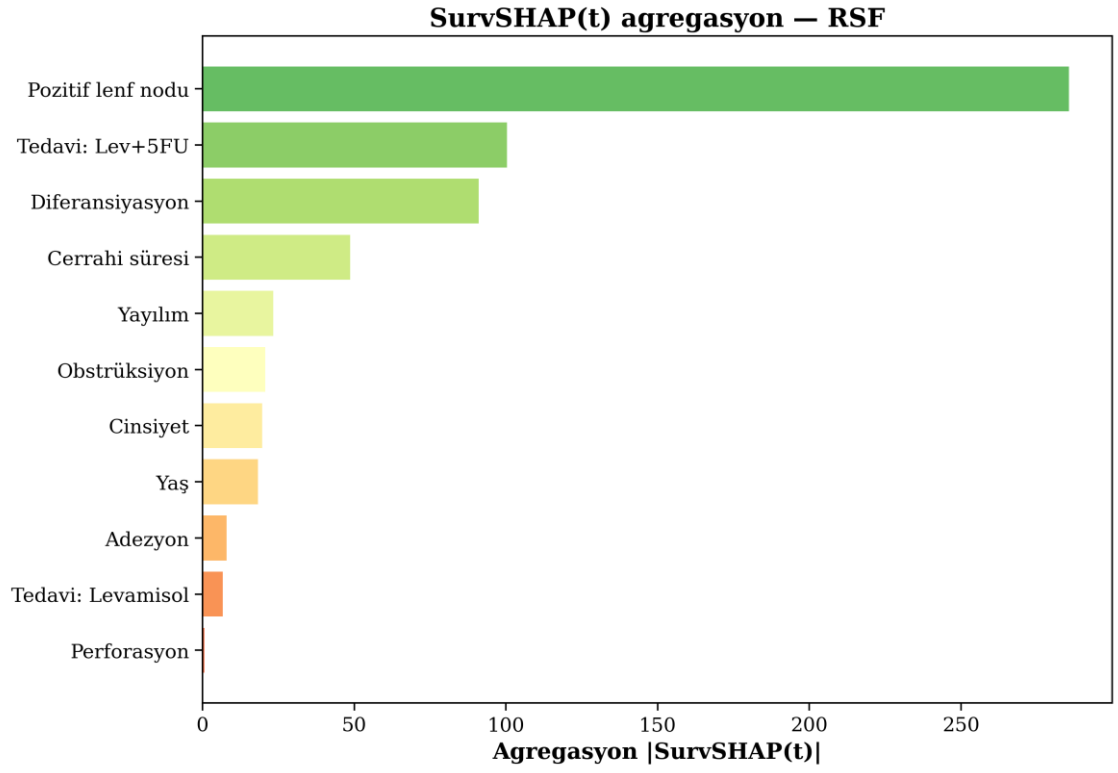
Şekil 4.26. DeepHit modeline ait global SurvSHAP(t) grafiği

Şekil 4.27'de CoxPH modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



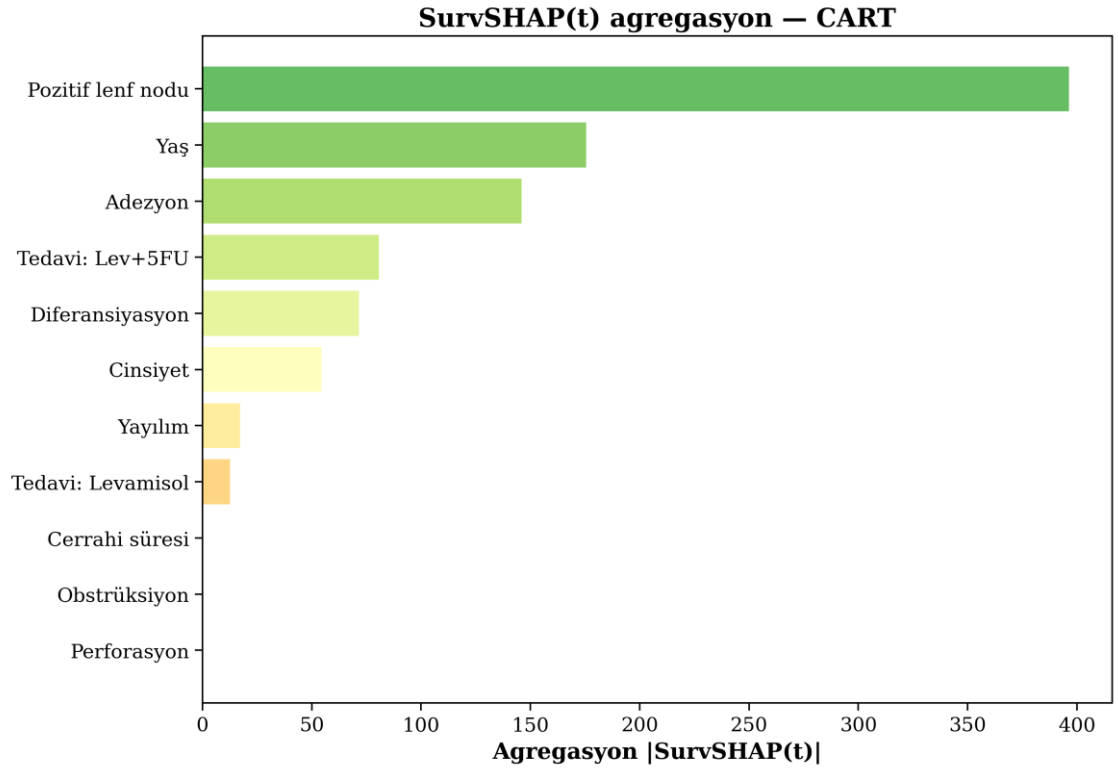
Şekil 4.27. CoxPH modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.28'de RSF modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



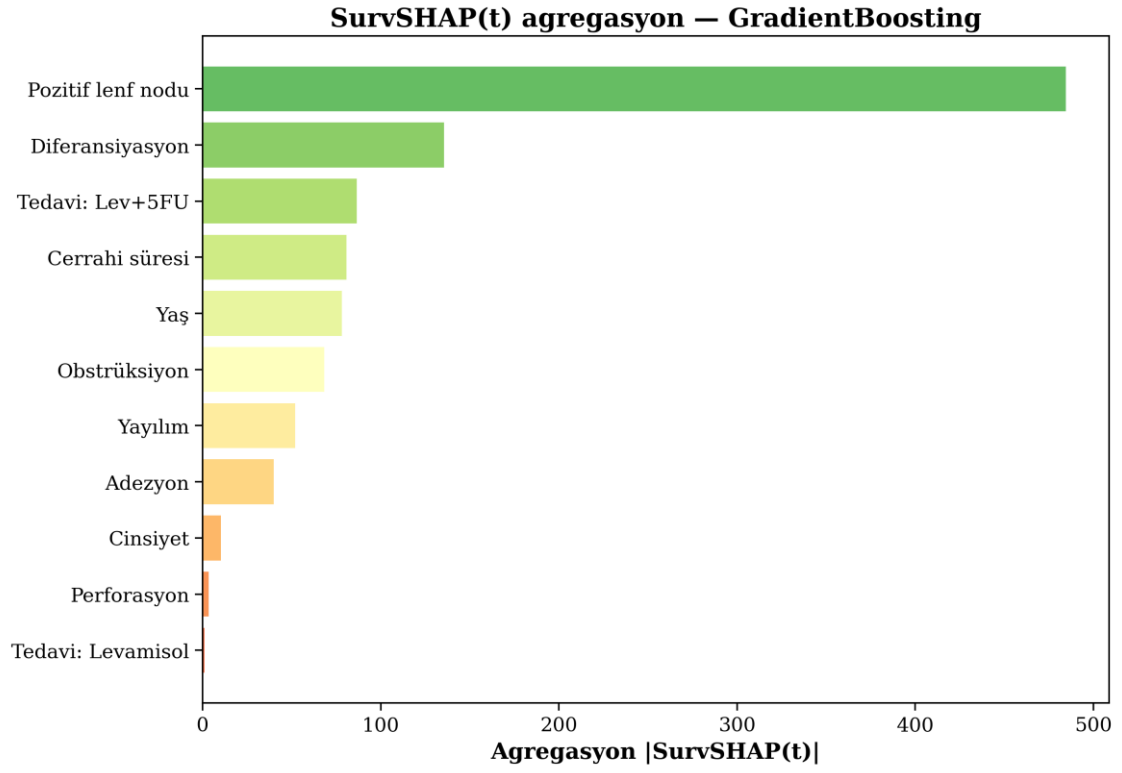
Şekil 4.28. RSF modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.29'da CART modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



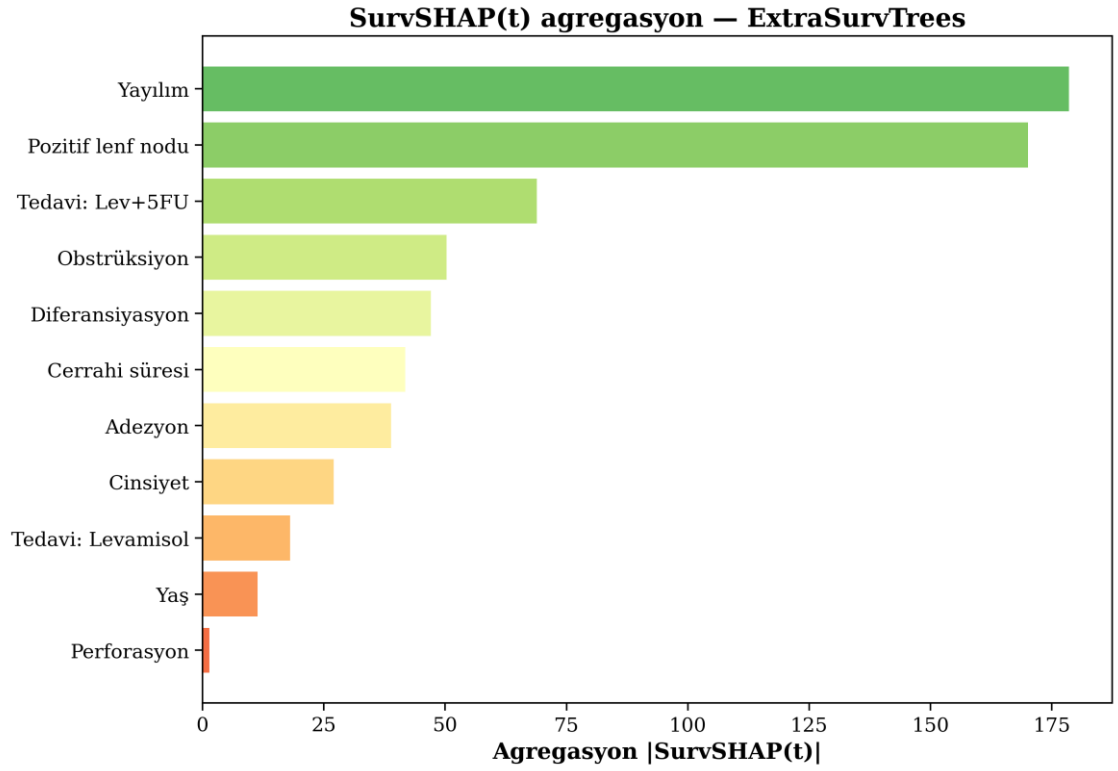
Şekil 4.29. CART modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.30'da GradientBoosting modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



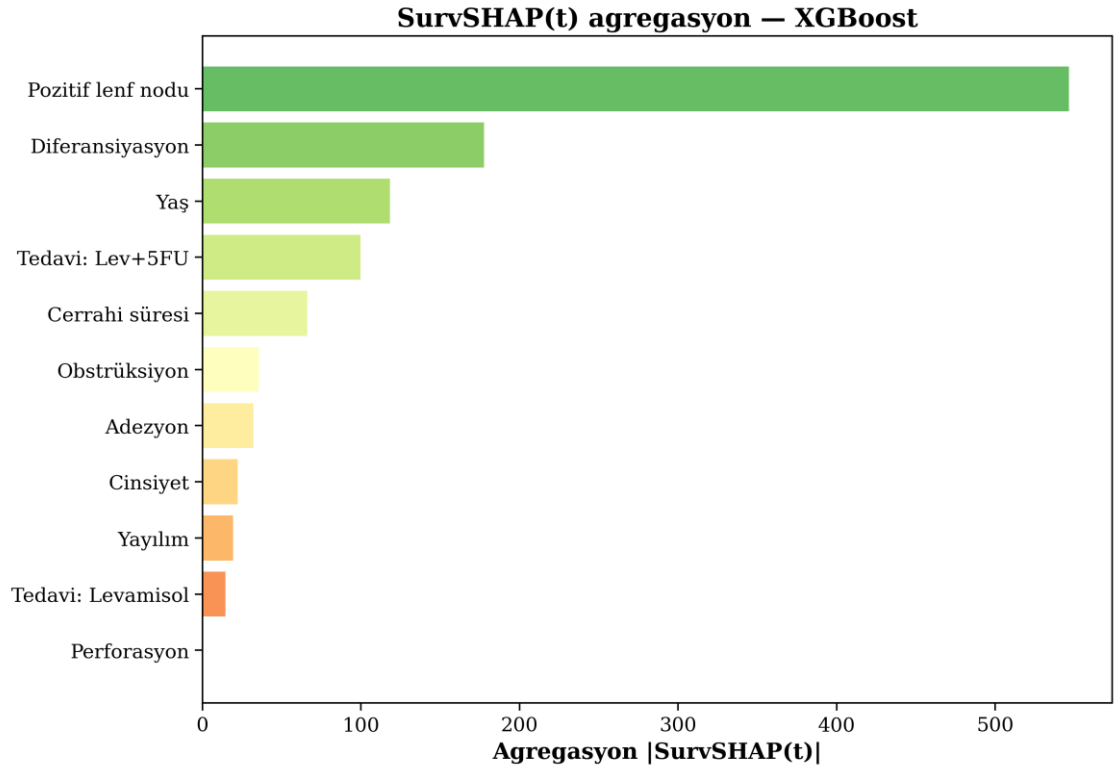
Şekil 4.30. GradientBoosting modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.31'de ExtraSurvTrees modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



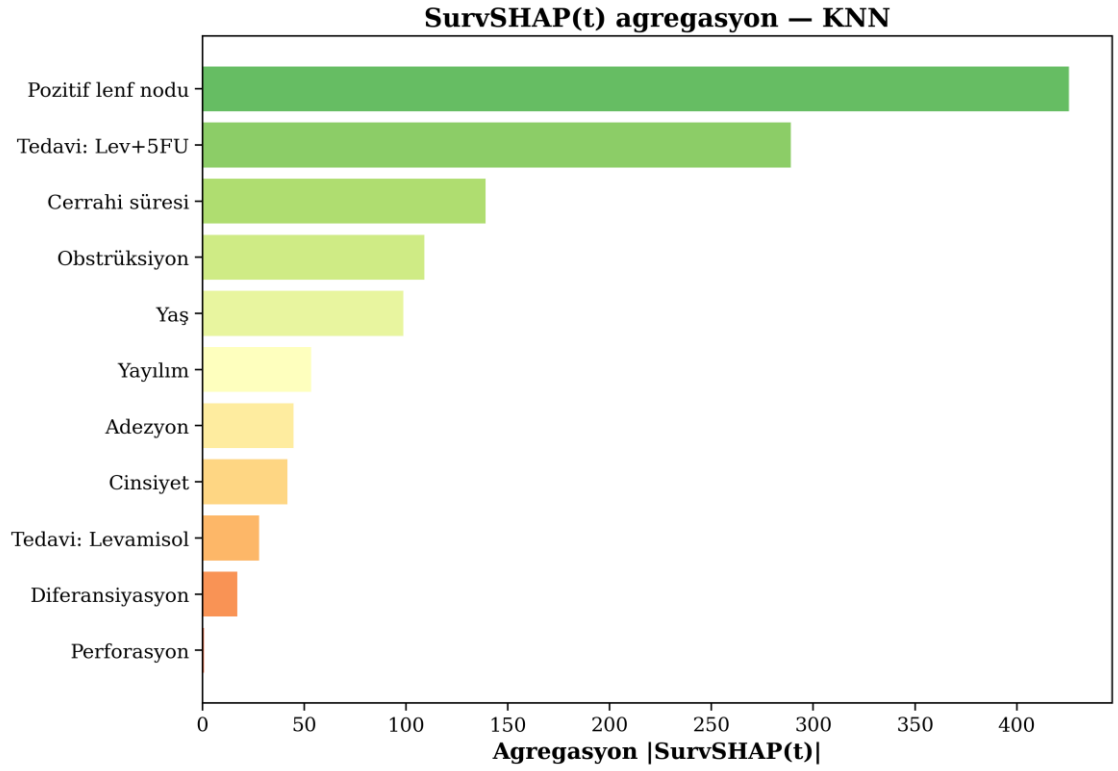
Şekil 4.31. ExtraSurvTrees modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.32'de XGBoost modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



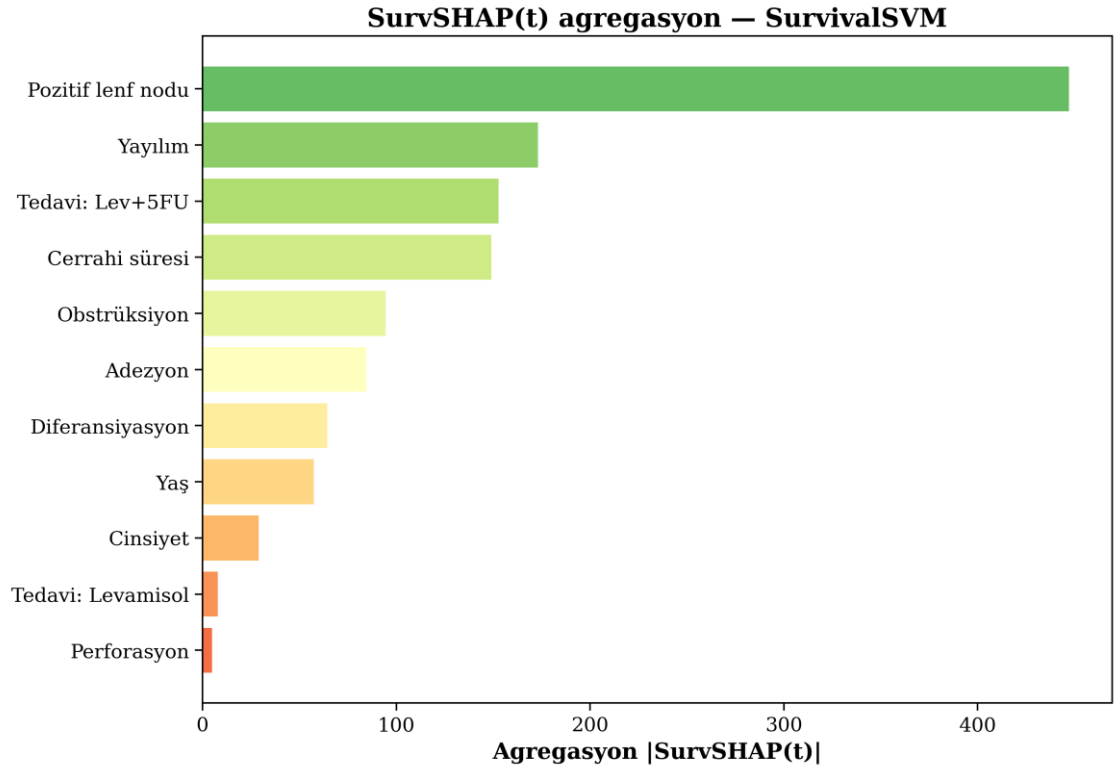
Şekil 4.32. XGBoost modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.33'te KNN modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



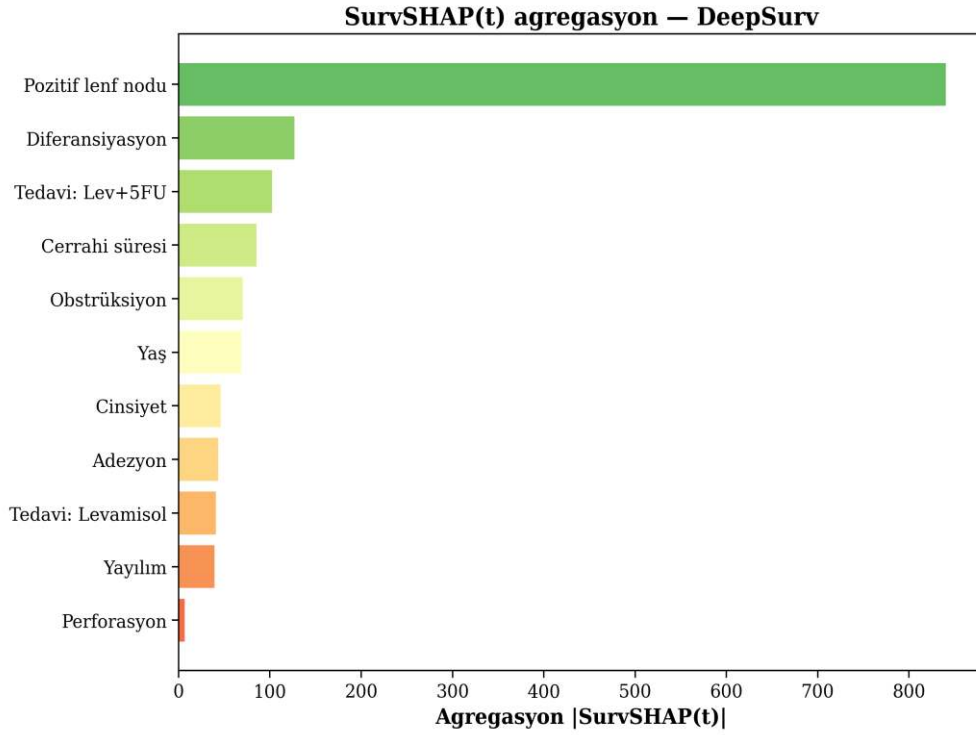
Şekil 4.33. KNN modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.34'te SurvivalSVM modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



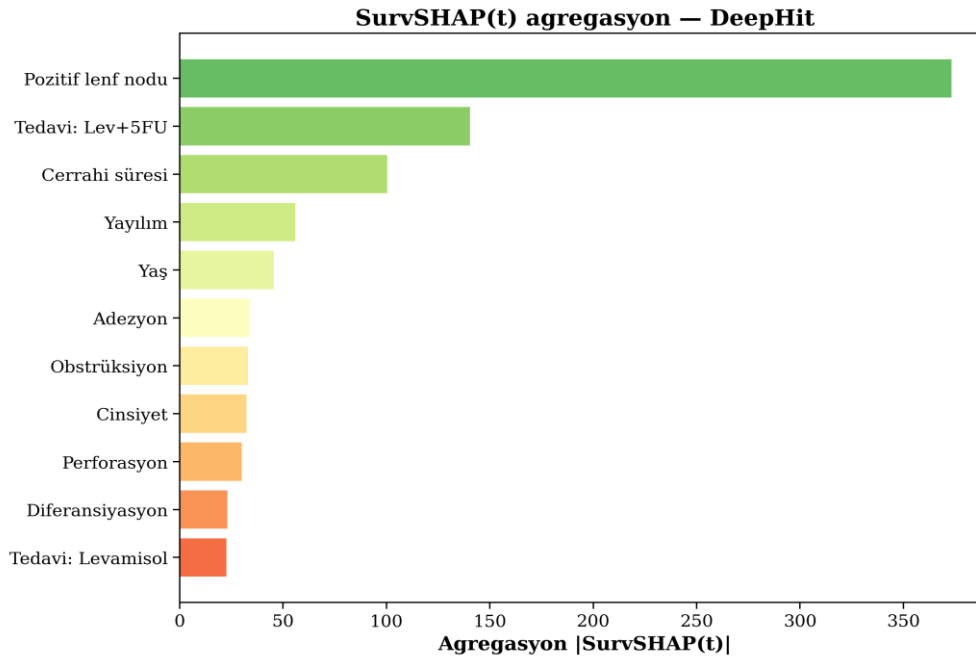
Şekil 4.34. SurvivalSVM modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.35'te DeepSurv modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı.



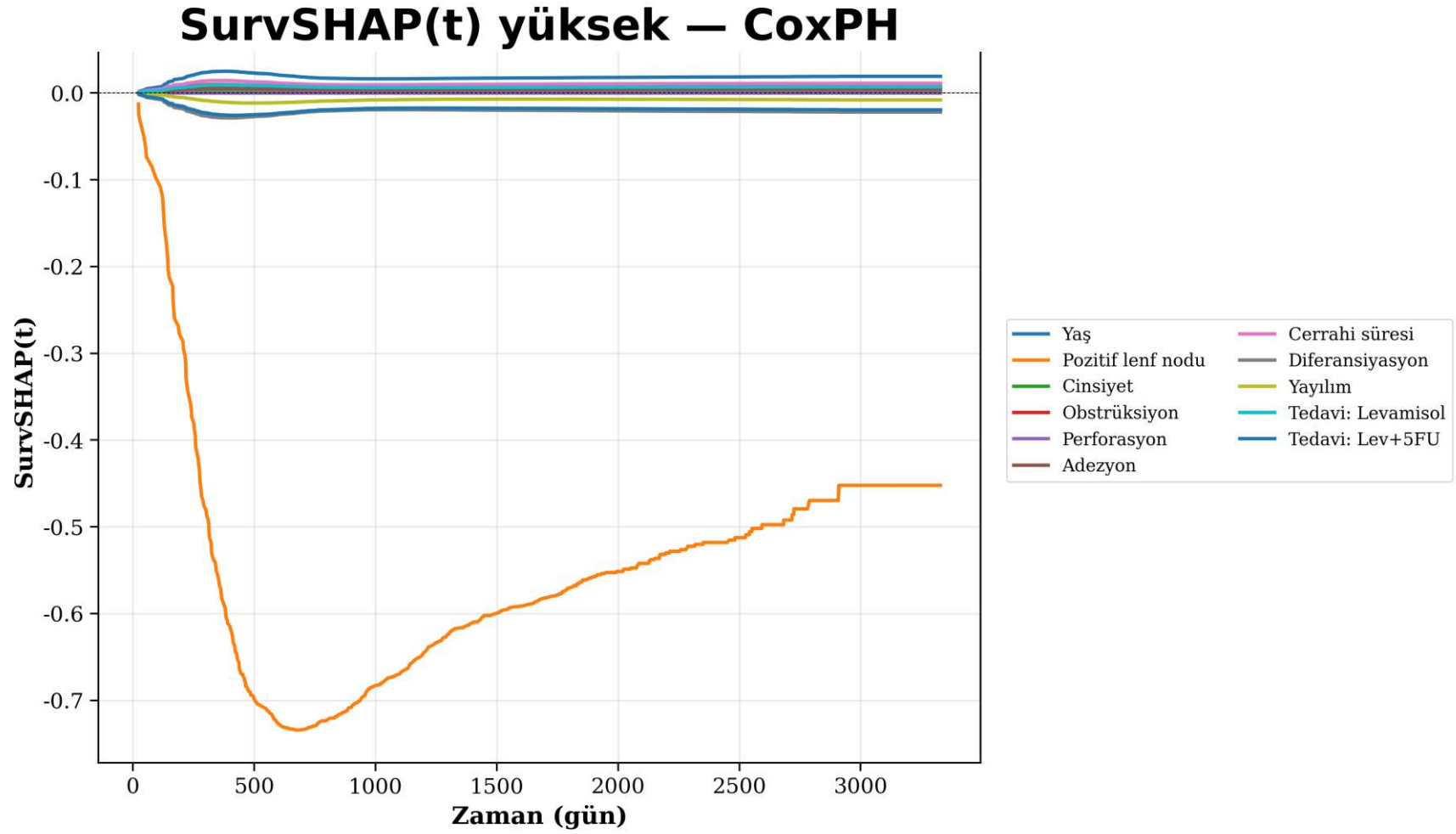
Şekil 4.35. DeepSurv modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.36'da DeepHit modeli için Tablo 4.3'deki agregasyon değerlerinin bar grafik gösterimi yer aldı. Her model için en yüksek ve en düşük riskli birer gözlem seçildi.



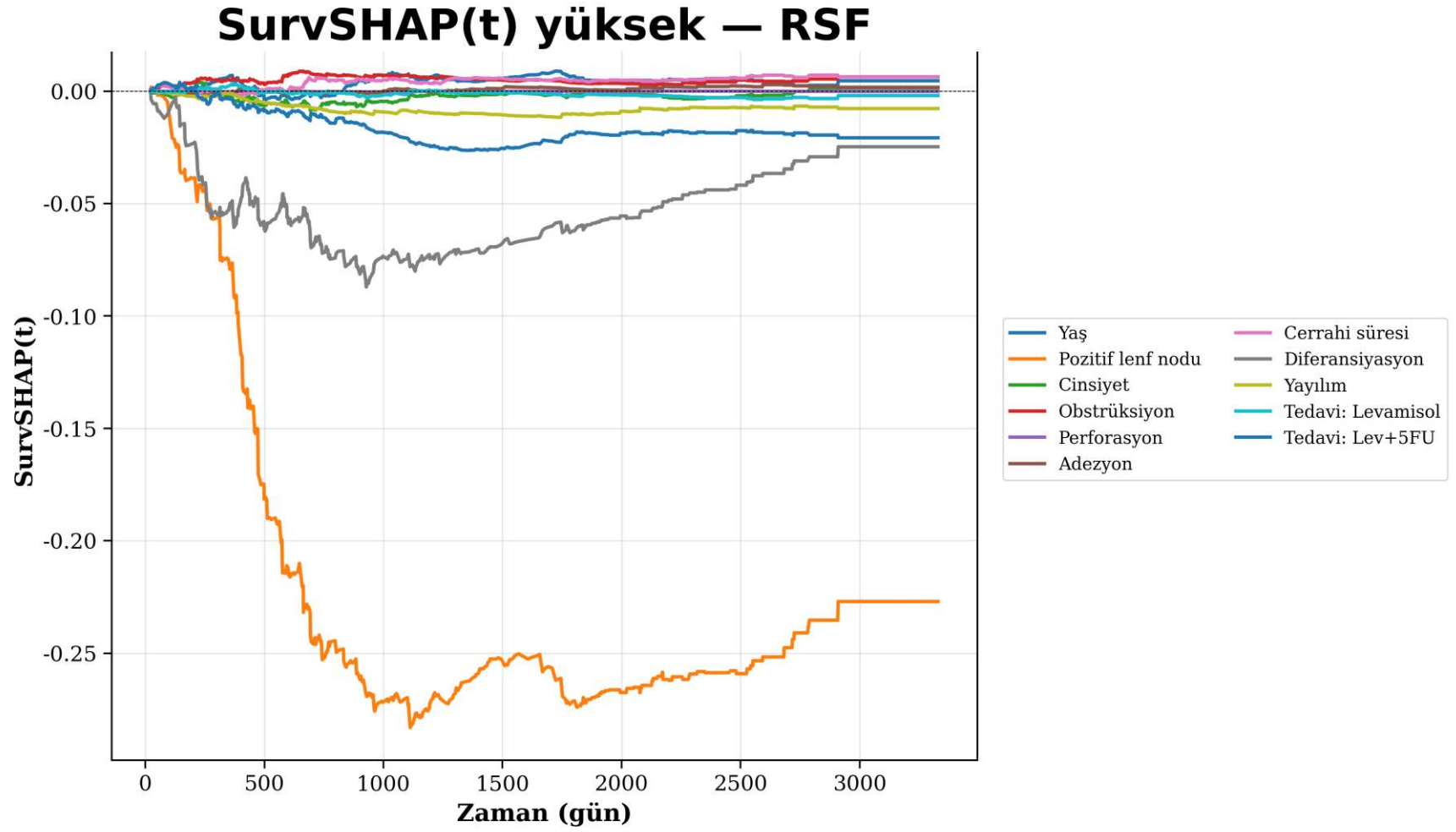
Şekil 4.36. DeepHit modeline ait SurvSHAP(t) agregasyon grafiği

Şekil 4.37'de CoxPH modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



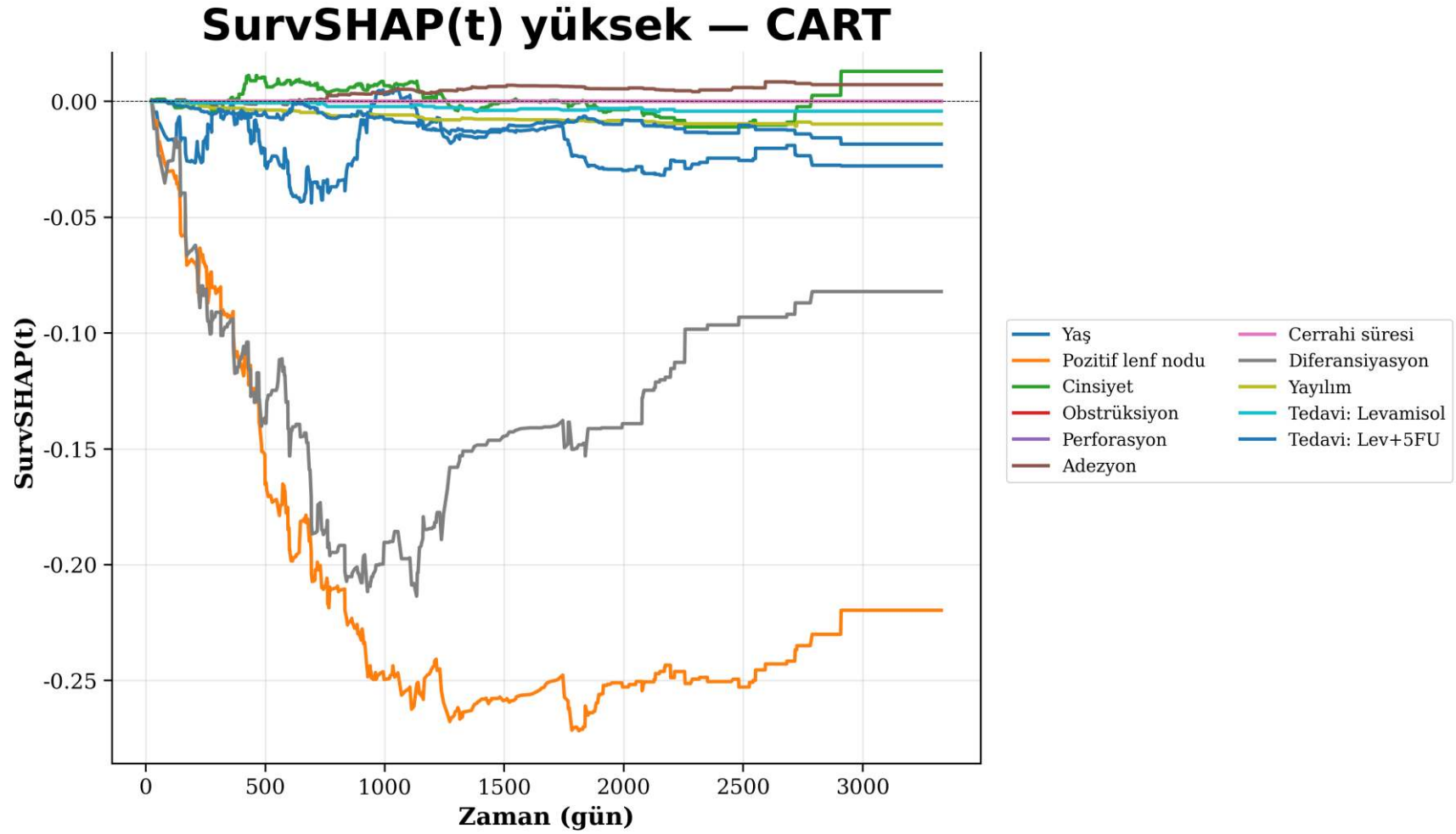
Şekil 4.37. CoxPH modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.38'de RSF modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



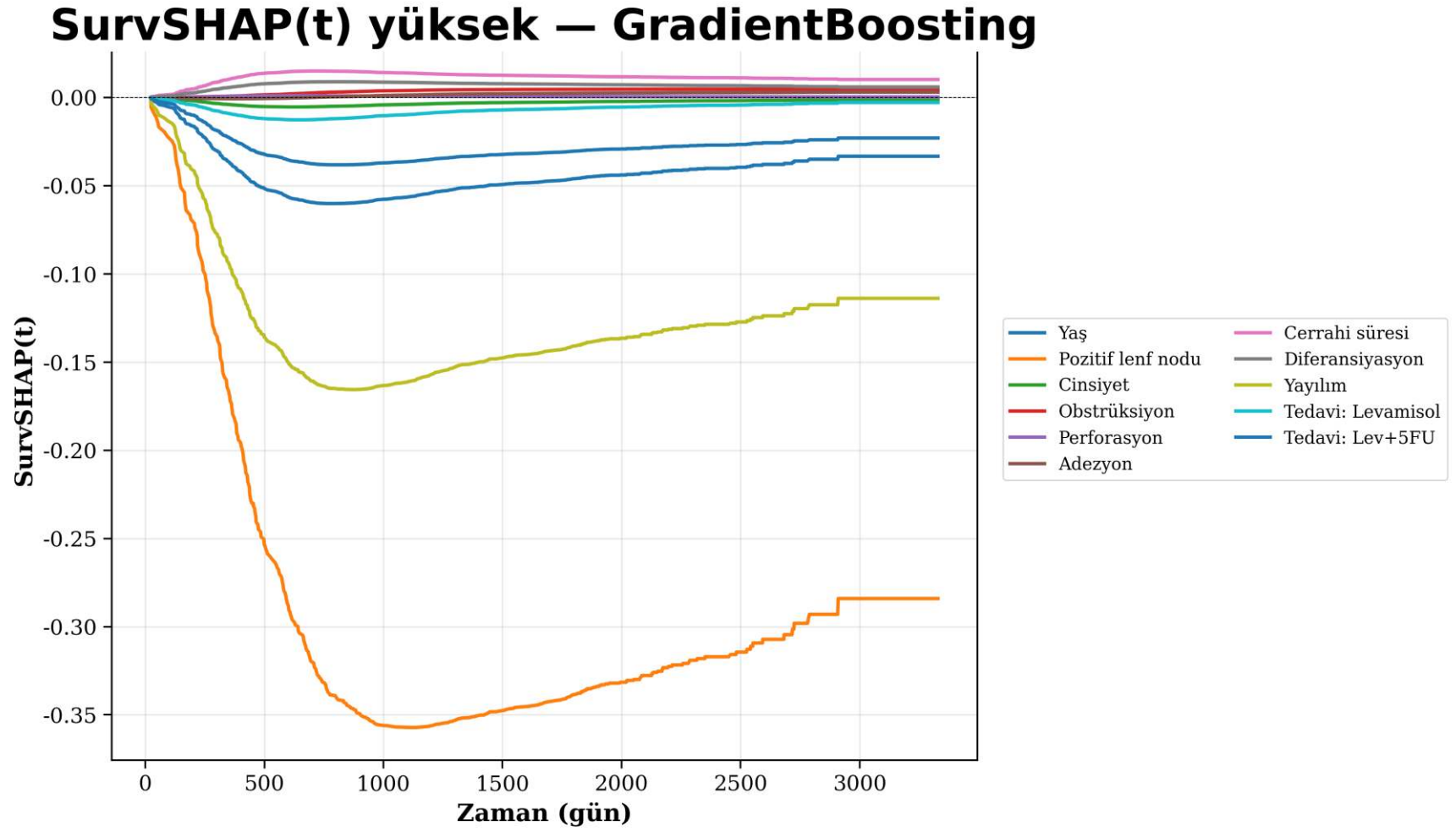
Şekil 4.38. RSF modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.39'da CART modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



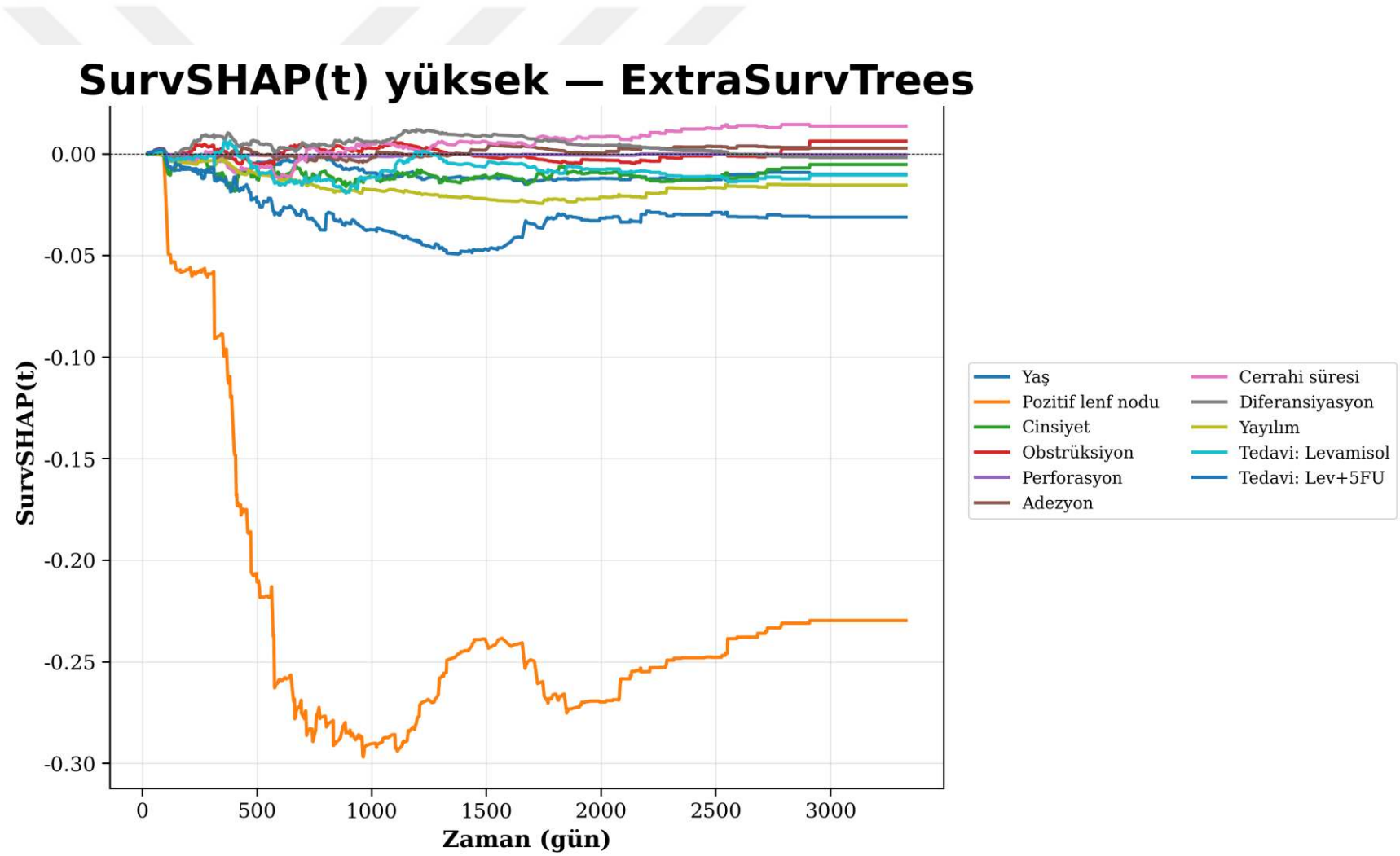
Şekil 4.39. CART modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.40'ta GradientBoosting modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



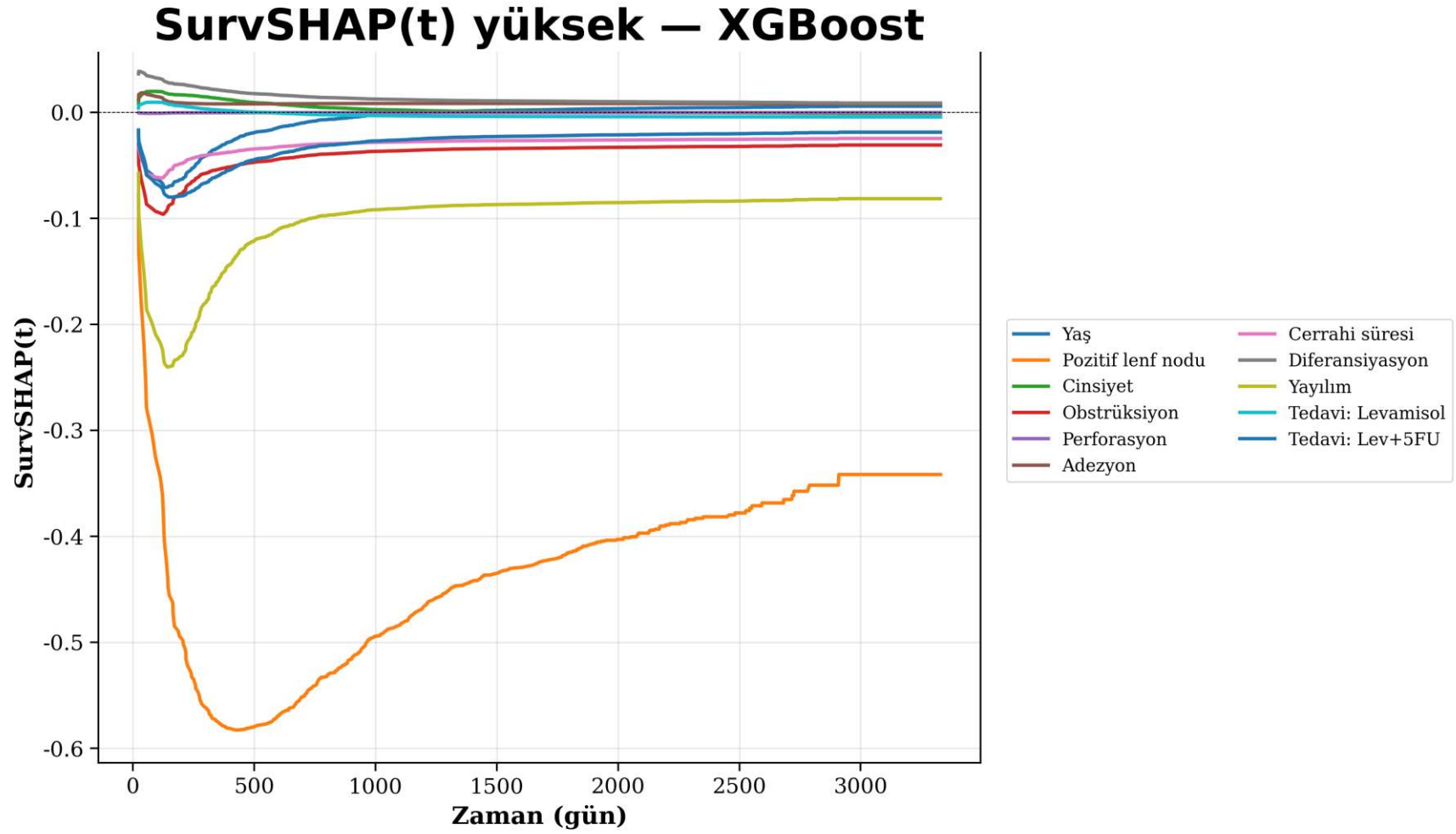
Şekil 4.40. GradientBoosting modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.41'de ExtraSurvTrees modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



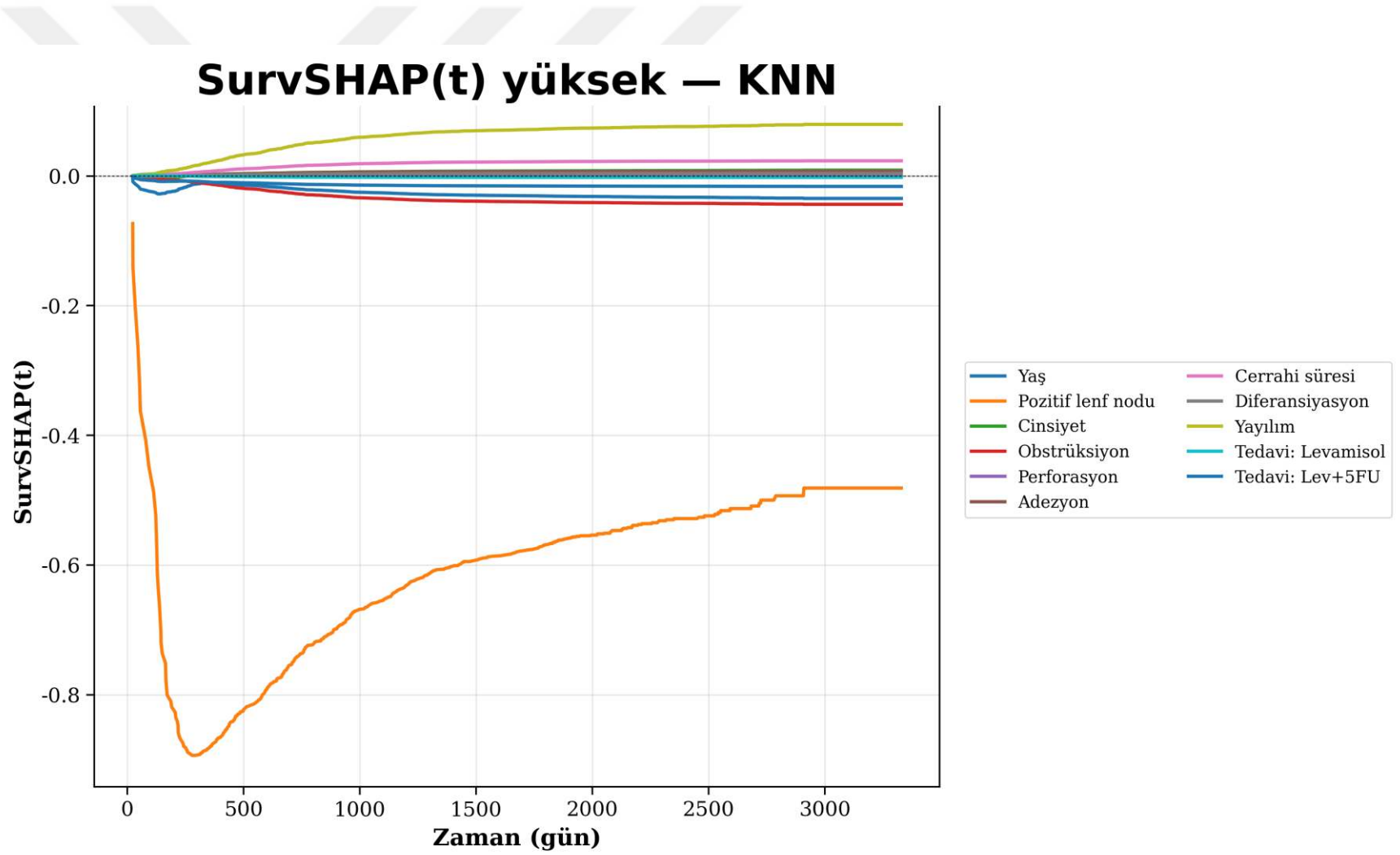
Şekil 4.41. ExtraSurvTrees modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.42'de XGBoost modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



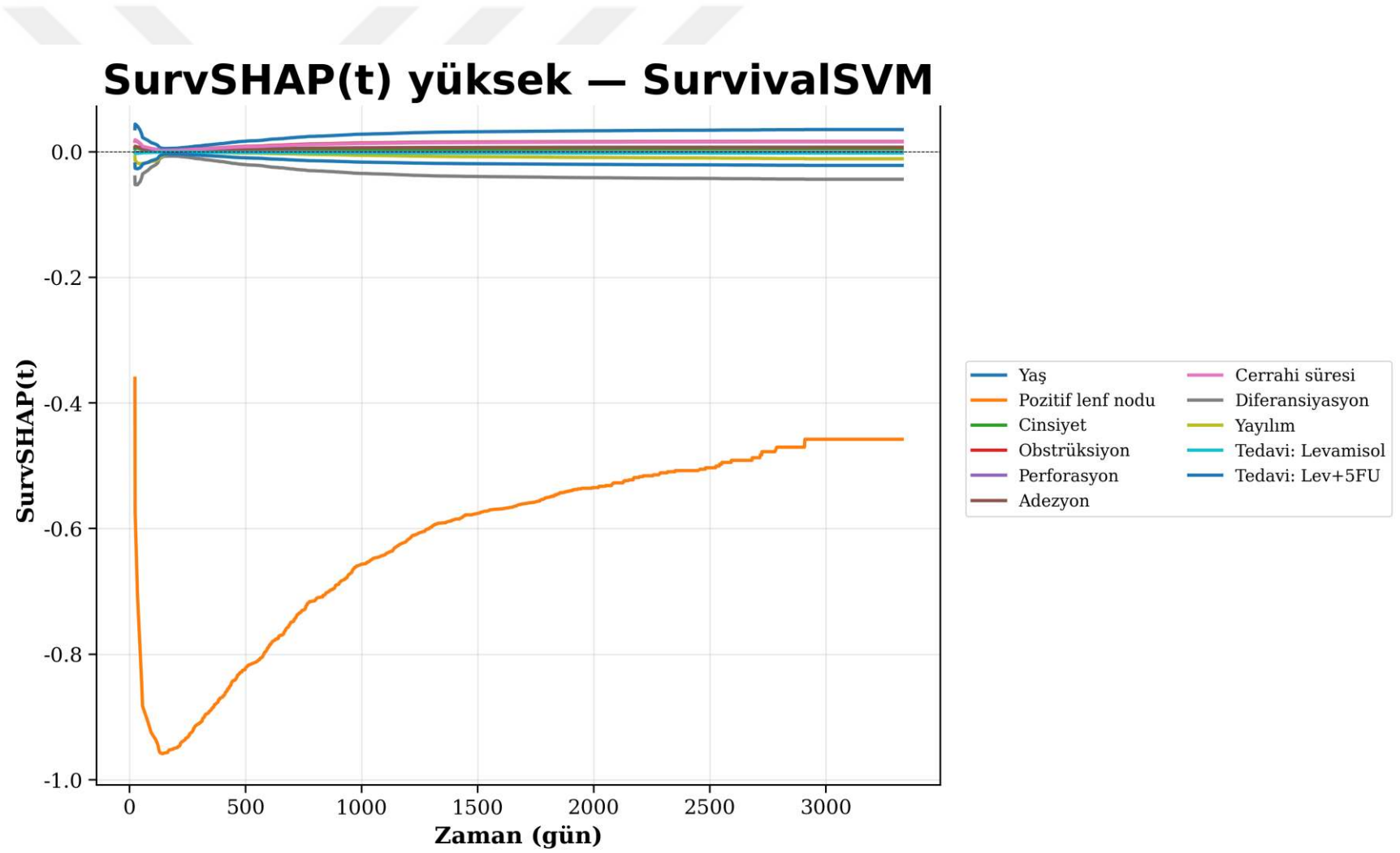
Şekil 4.42. XGBoost modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.43'te KNN modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



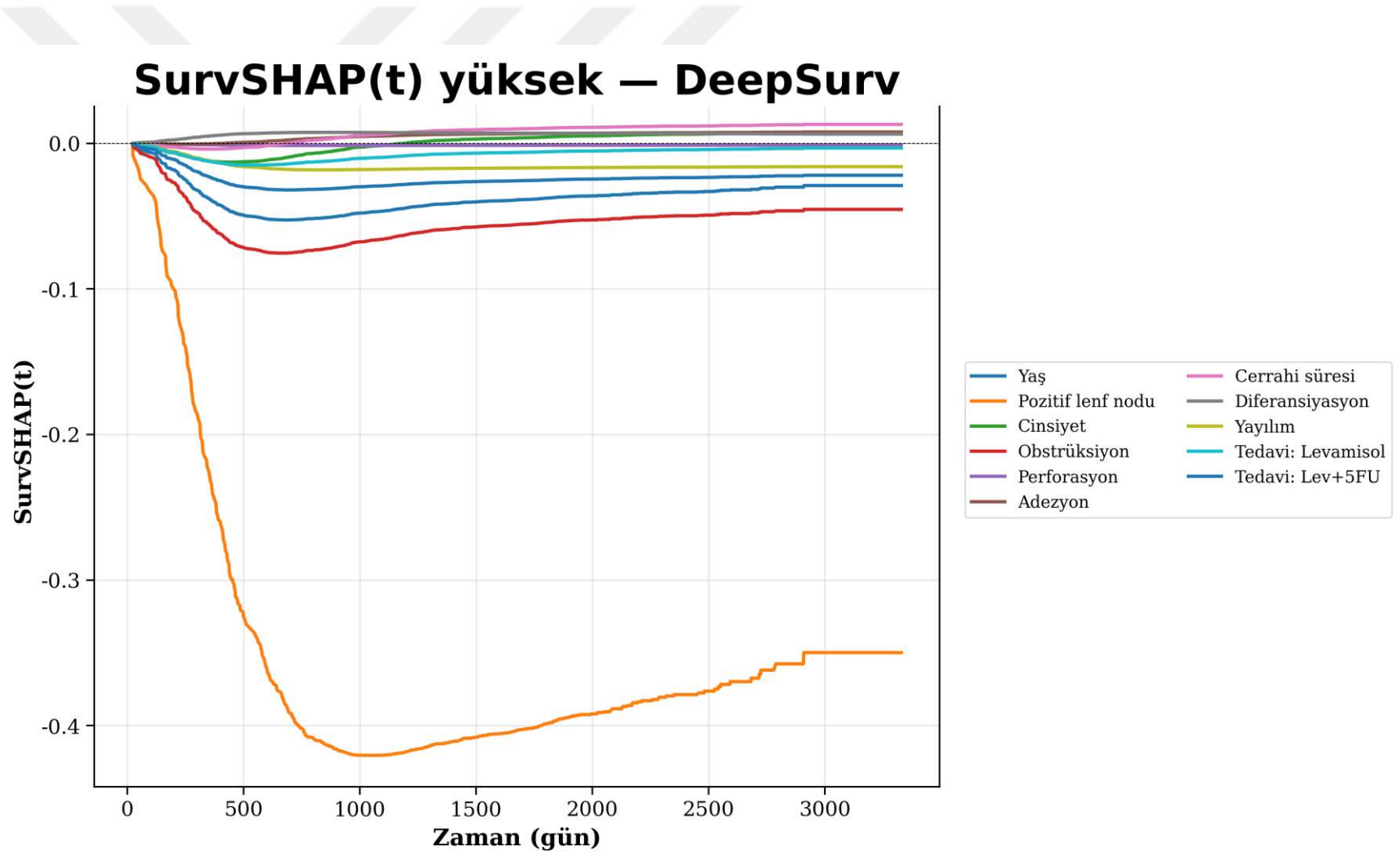
Şekil 4.43. KNN modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.44'te SurvivalSVM modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



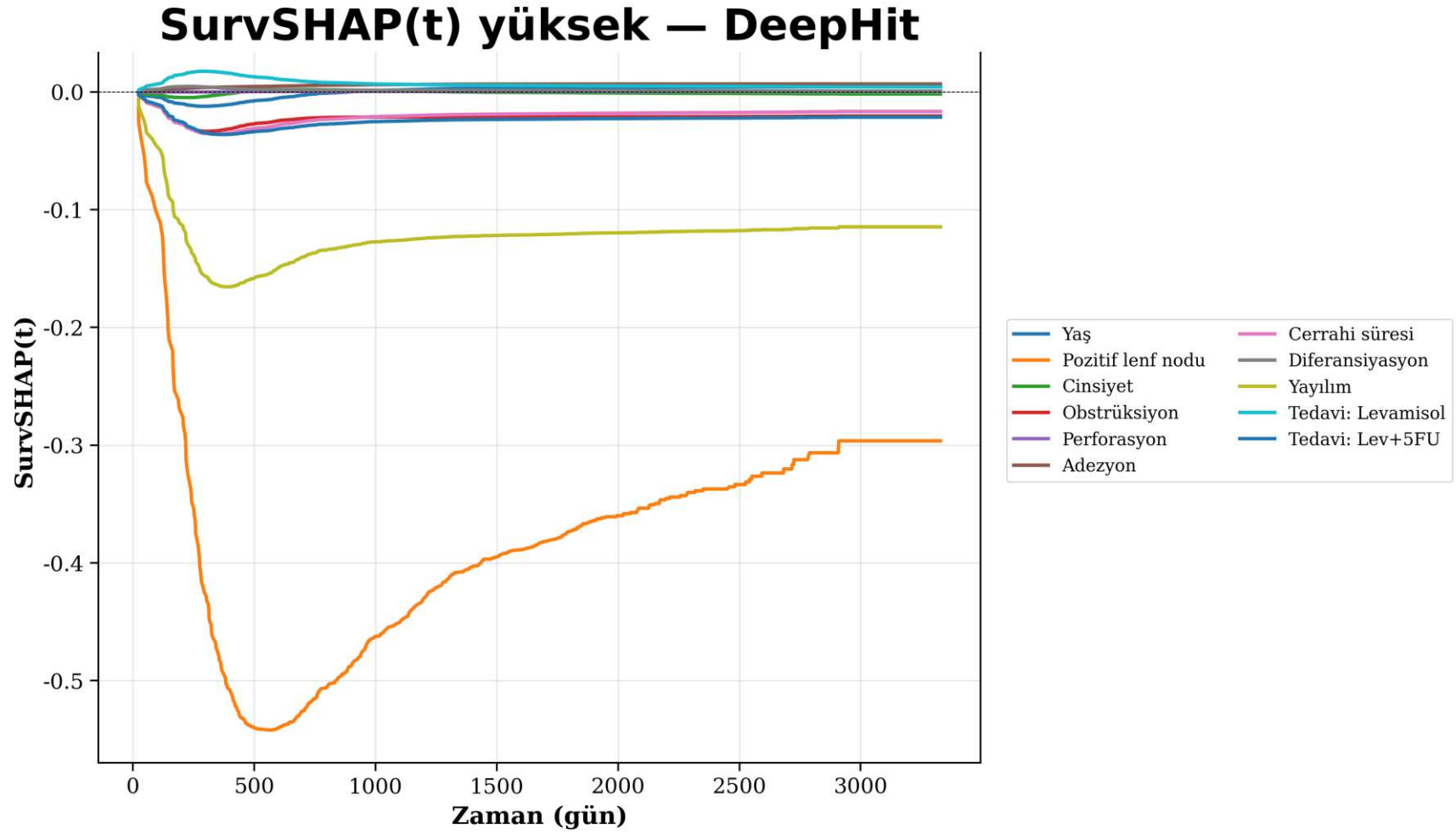
Şekil 4.44. SurvivalSVM modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.45'te DeepSurv modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



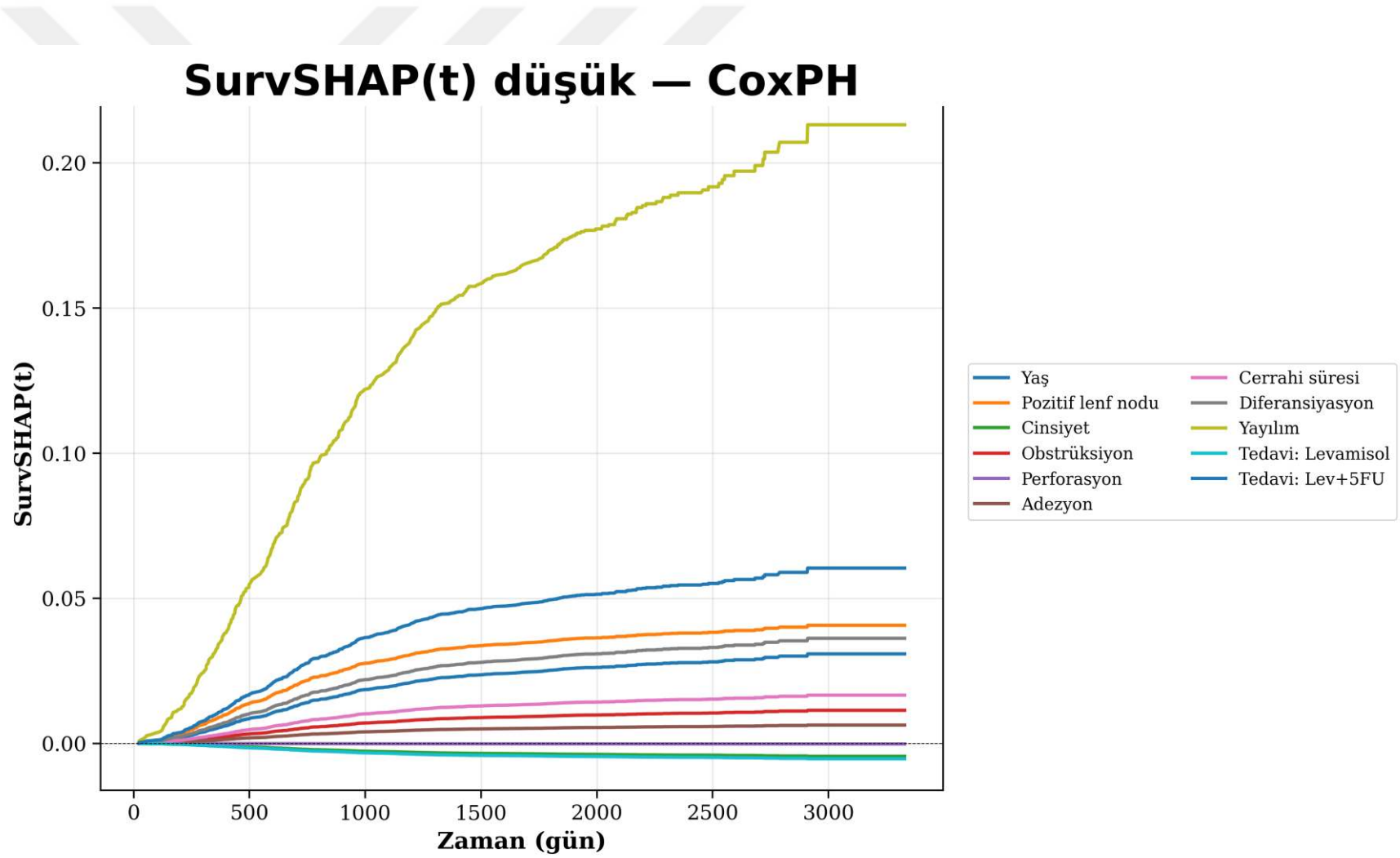
Şekil 4.45. DeepSurv modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.46'da DeepHit modelinin en yüksek riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı negatif yönde (sıfırın altında) etki gösterdi.



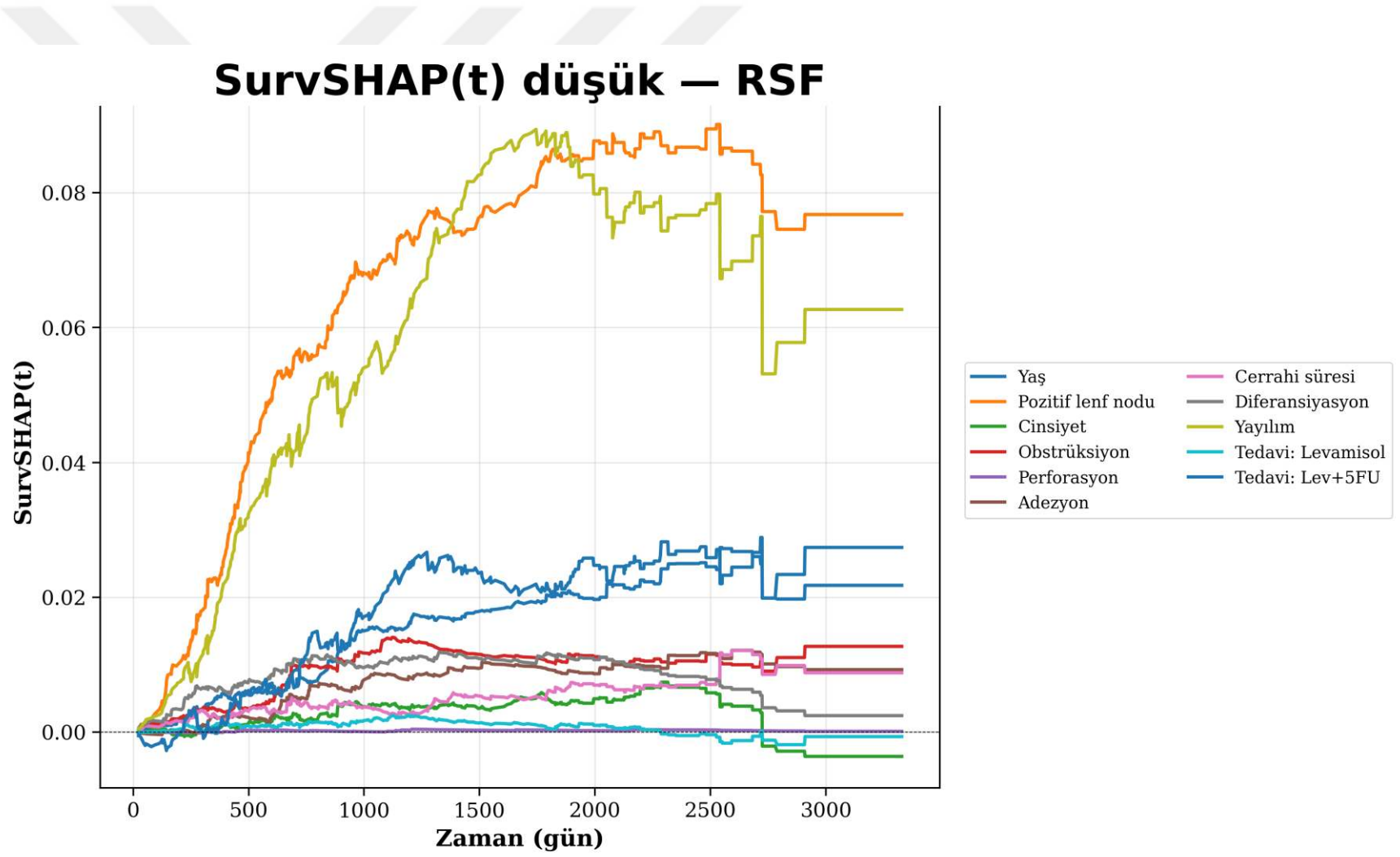
Şekil 4.46. DeepHit modeline ait SurvSHAP(t) yüksek riskli hasta profili

Şekil 4.47'de CoxPH modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



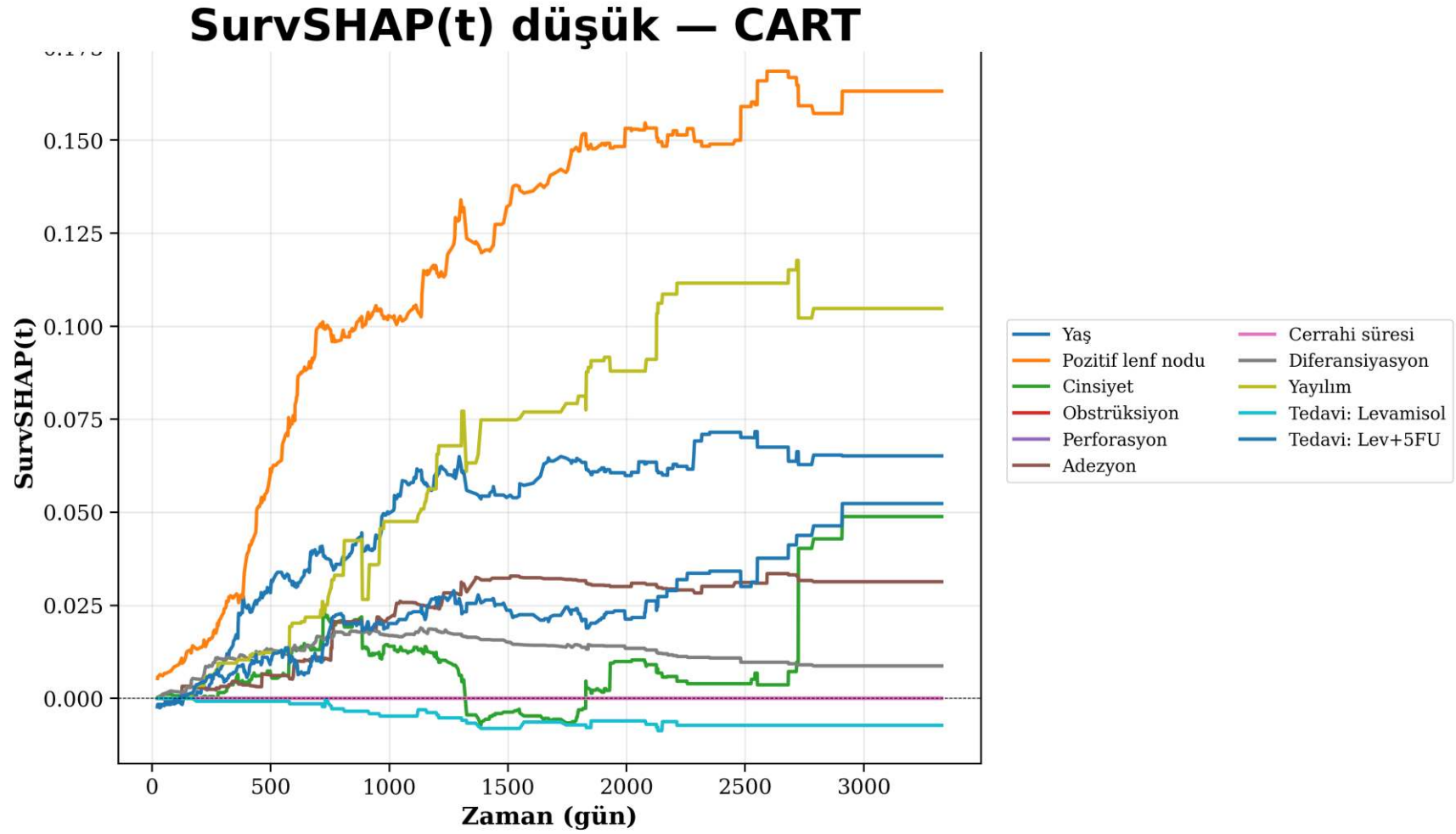
Şekil 4.47. CoxPH modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.48'de RSF modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



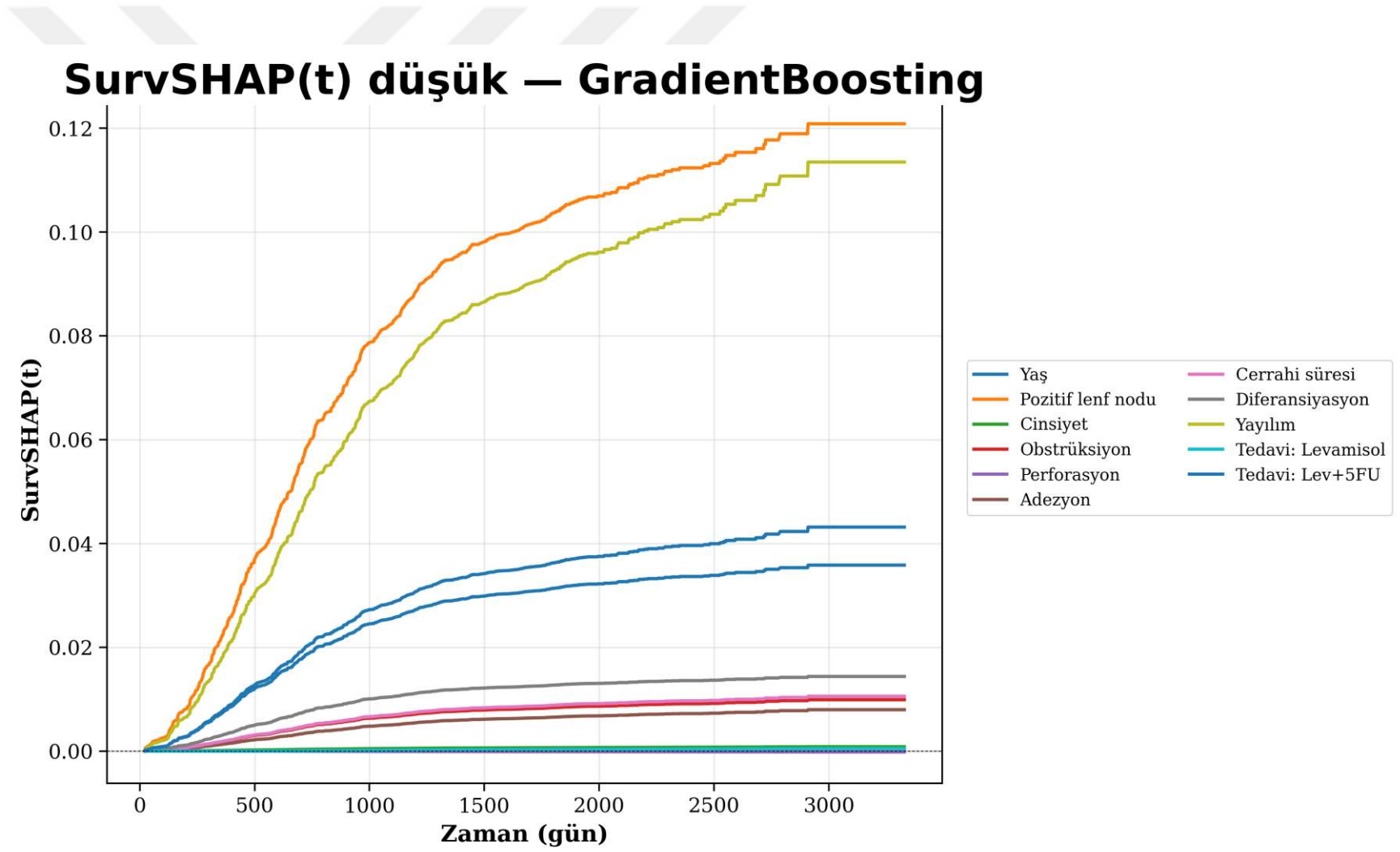
Şekil 4.48. RSF modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.49'da CART modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı ve yayılım pozitif yönde (sıfırın üzerinde) etki gösterdi.



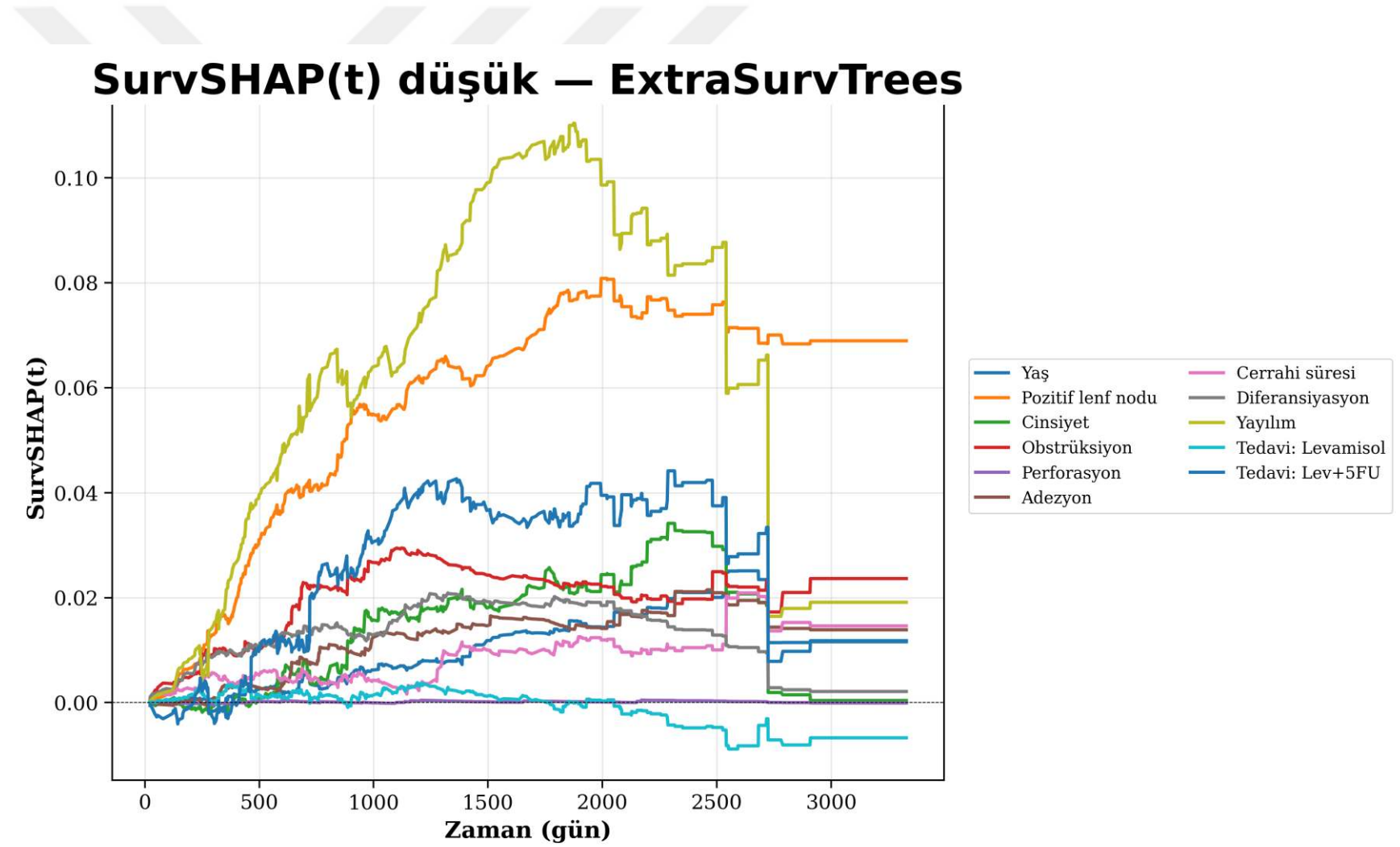
Şekil 4.49. CART modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.50'de GradientBoosting modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



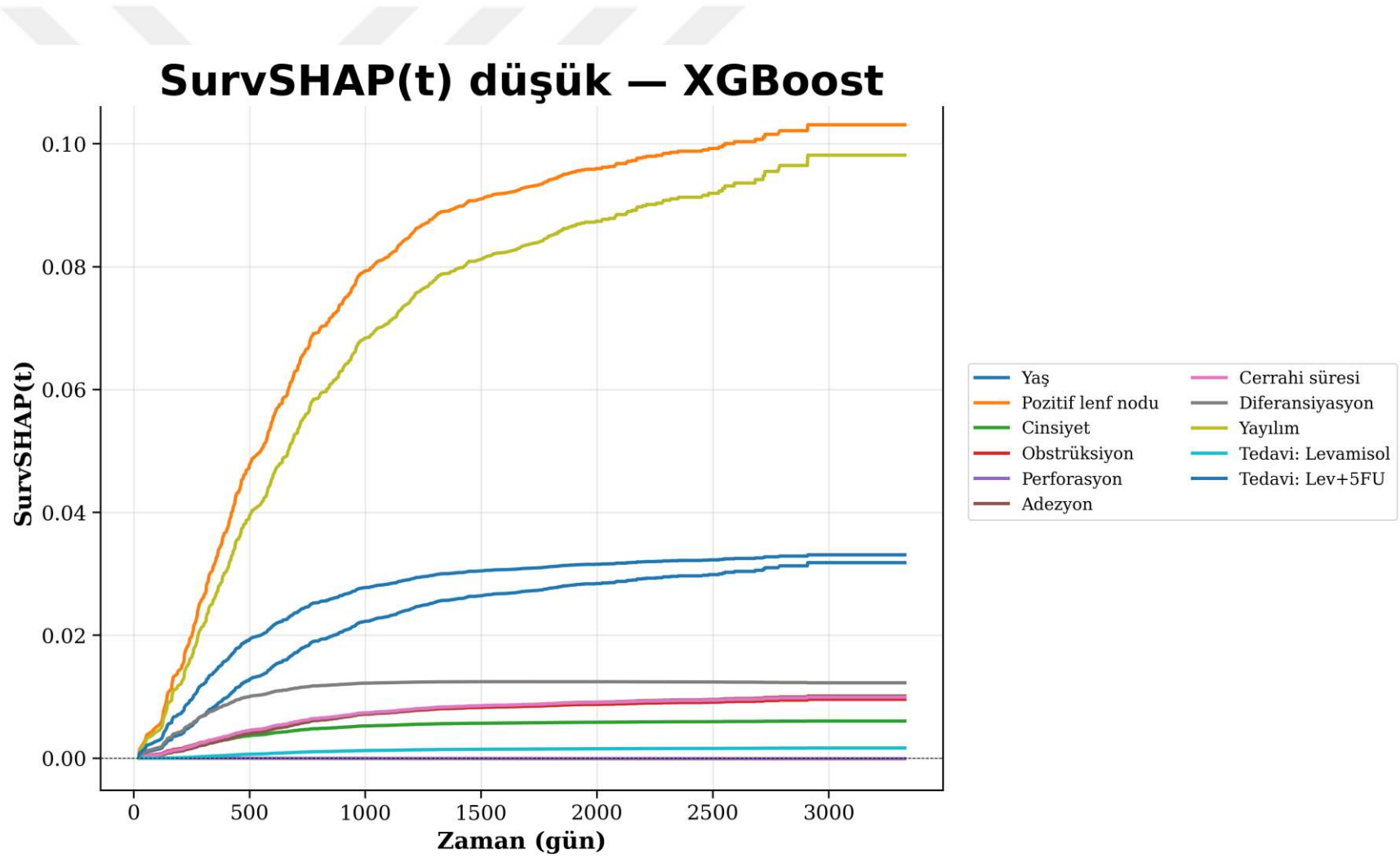
Şekil 4.50. GradientBoosting modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.51'de ExtraSurvTrees modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



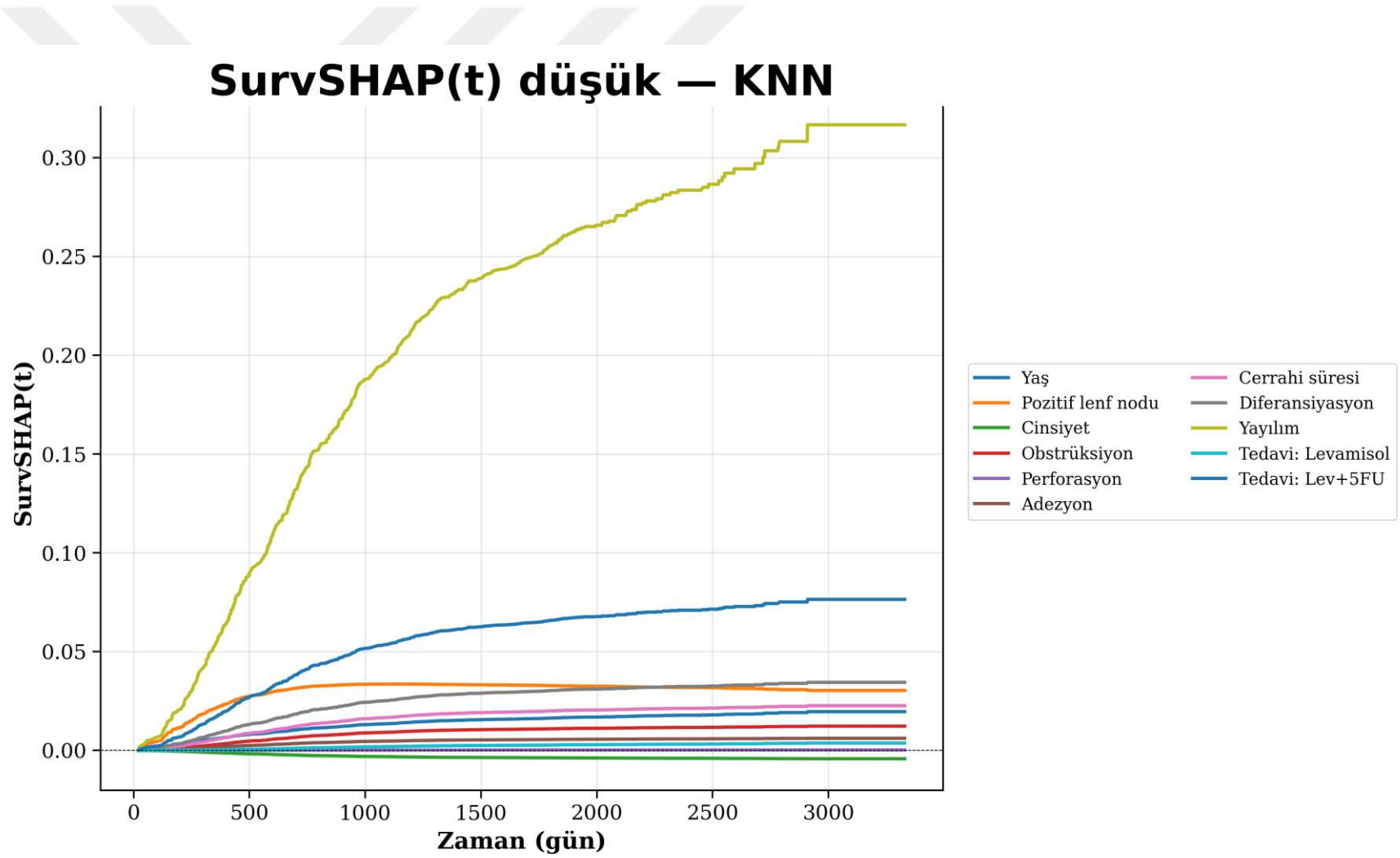
Şekil 4.51. ExtraSurvTrees modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.52'de XGBoost modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



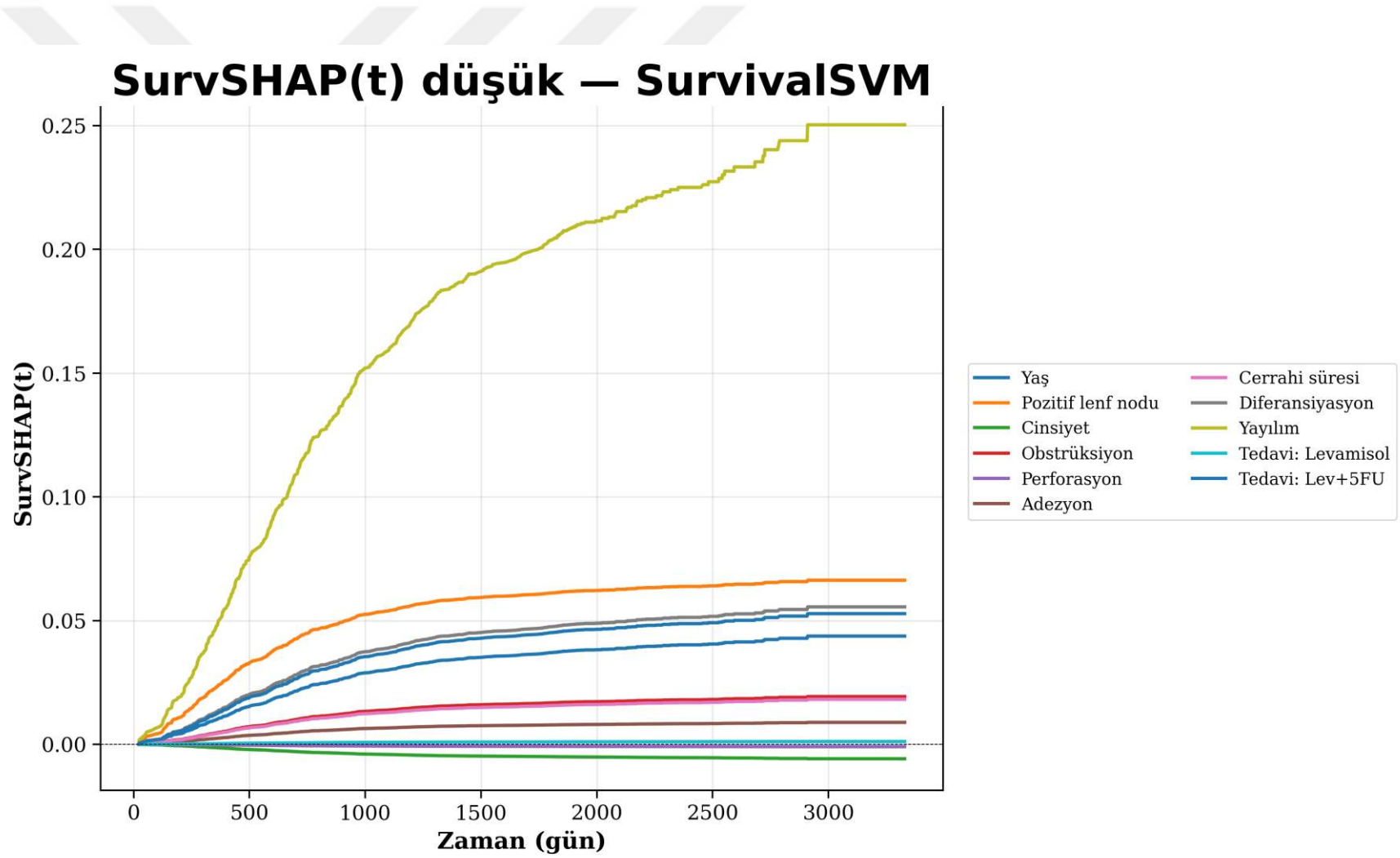
Şekil 4.52. XGBoost modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.53'te KNN modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



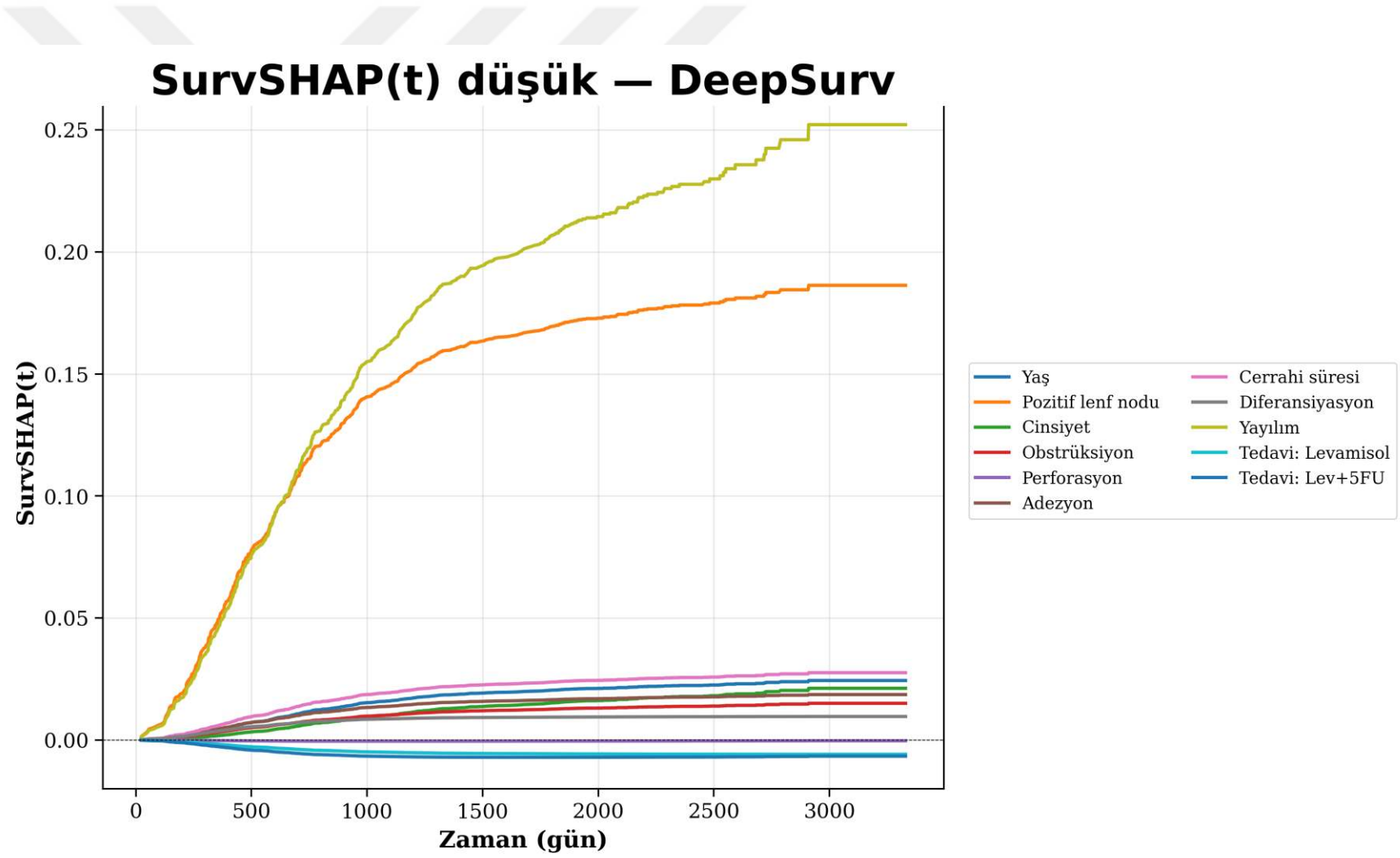
Şekil 4.53. KNN modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.54'te SurvivalSVM modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



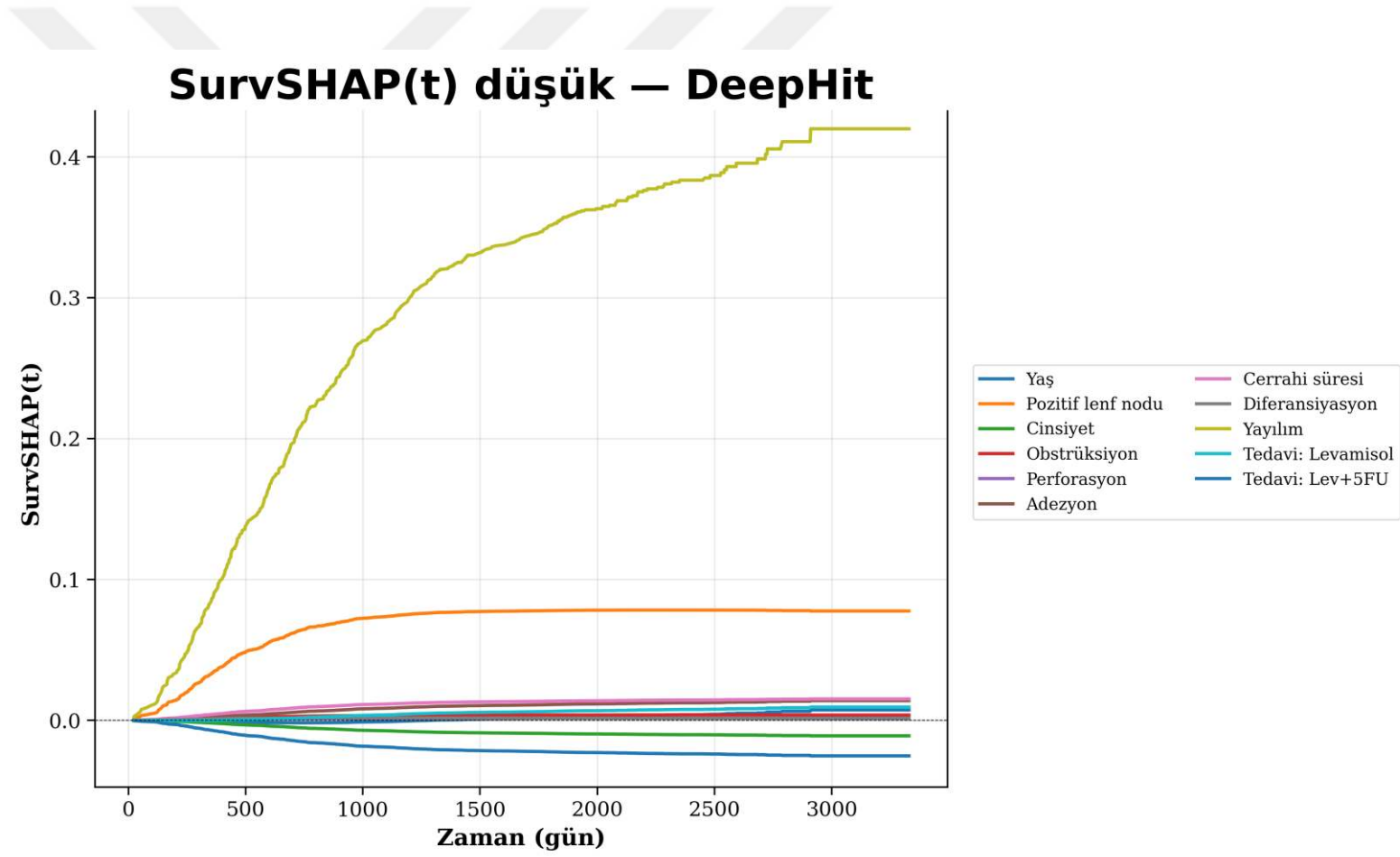
Şekil 4.54. SurvivalSVM modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.55'te DeepSurv modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



Şekil 4.55. DeepSurv modeline ait SurvSHAP(t) düşük riskli hasta profili

Şekil 4.56'da DeepHit modelinin en düşük riskli hasta profili yer aldı. Pozitif lenf düğümü sayısı pozitif yönde (sıfırın üzerinde) etki gösterdi.



Şekil 4.56. DeepHit modeline ait SurvSHAP(t) düşük riskli hasta profili

Açıklanabilir sağkalım analizinin üçüncü katmanı olarak SurvLIME yöntemi uygulandı. tarafından önerilen bu yaklaşım, karmaşık sağkalım modellerinin bireysel düzeydeki tahminlerini Cox orantılı tehlike modeli çerçevesinde lokal olarak yaklaştırdı (19). Böylece her bir hasta için hangi klinik özniteliklerin sağkalım tahminini ne yönde etkilediği sayısal katsayılarla ifade edilebildi. Analizde R survex paketi kullanıldı (91); CoxPH modelinde beş, RSF modelinde dört farklı risk düzeyindeki hasta profili incelendi.

Değerlendirme için risk skoru dağılımının farklı bölgelerinden beş gözlem seçildi: en yüksek risk, yüksek risk (90. persentil), orta risk (medyan), düşük risk (10. persentil) ve en düşük risk. Bu seçim stratejisi, SurvLIME katsayılarının risk spektrumu boyunca nasıl değiştiğini sistematik biçimde ortaya koymayı amaçladı (Tablo 4.5).

Tablo 4.5. Kullanılan modeller ve parametreleri

Model	En İyi Parametreler
CoxPH	alpha=1.0
KNN	n_neighbors=15, weights=uniform, metric=manhattan
SVM	alpha=0.001, rank_ratio=1.0, max_iter=50
RSF	n_estimators=300, max_depth=5, min_split=12, min_leaf=2
CART	max_depth=15, min_split=30, min_leaf=8
GBSurv	n_estimators=300, lr=0.018, max_depth=2, subsample=0.693
ExtraTrees	n_estimators=300, max_depth=15, min_split=29, min_leaf=2
XGBoost	n_estimators=412, max_depth=2, lr=0.011, subsample=0.607
DeepSurv	layers=2, nodes=32, dropout=0.378, lr=0.001, epochs=100
DeepHit	layers=3, nodes=128, dropout=0.414, lr=0.0002, epochs=100

Tablo 4.6 incelendiğinde, en yüksek riskli hastada Yayılım: Bitişik yapılar özniteliğinin en büyük pozitif katsayıya (0.6491) sahip olduğu; buna karşılık Yayılım: Kas tabakası (-0.4934) ve Lev+5FU tedavisi (-0.4429) özniteliklerinin güçlü negatif (koruyucu) katsayılar sergilediği görüldü. Adezyon: Evet (0.3259) ve Obstrüksiyon: Evet (0.2657) ek pozitif katkı sağlarken Diferansiyasyon: Orta (-0.2007) negatif yönde etkilidir. Bu örüntü, bitişik yapılara invazyonun yerel Cox yaklaşıklamasında sağkalım riskini artıran; etkili kemoterapi kombinasyonunun ise riski azaltan başat faktörler olduğuna işaret etti. Risk düzeyi azaldıkça (yüksek, orta, düşük ve en düşük riskli hastalarda) katsayı büyüklükleri belirgin biçimde küçülmekte ve sifıra yaklaştı; bu da SurvLIME yerel katsayılarının hasta risk profiline duyarlı olduğunu ortaya koydu.

Buradaki değerlerin yerel Cox modelinin ham katsayıları (β) olduğu, ilgili şekillerde ise survex paketinin yerel önem (local importance) ölçütünün gösterildiği unutulmamalıdır.

Tablo 4.6. SurvLIME katsayıları: CoxPH modeli (5 gözlem)

Öznitelik	En Yüksek	Yüksek	Orta	Düşük	En Düşük
Yayılım: Bitişik yapılar	0.6491	0.0025	0.0055	0.0335	0.0050
Yayılım: Kas tabakası	-0.4934	-0.0028	-0.0014	-0.0023	-0.0029
Tedavi: Lev+5FU	-0.4429	-0.0035	-0.0019	-0.0003	-0.0001
Adezyon: Evet	0.3259	0.0009	0.0017	-0.0005	0.0030
Perforasyon: Evet	-0.2607	0.0048	-0.0003	0.0073	0.0032
Obstrüksiyon: Evet	0.2657	0.0015	0.0019	0.0006	0.0003
Diferansiyasyon: Orta	-0.2007	-0.0033	-0.0009	-0.0024	-0.0014
Yayılım: Seroza	0.1381	-0.0004	-0.0010	-0.0005	-0.0020
Cerrahi süresi: Uzun	0.1391	0.0000	0.0011	0.0008	-0.0002
Tedavi: Levamisol	-0.0329	-0.0019	-0.0006	0.0003	0.0012
Cinsiyet: Erkek	0.0310	0.0004	-0.0009	-0.0011	-0.0004
Pozitif lenf nodu	0.0232	0.0005	0.0001	0.0003	0.0002
Diferansiyasyon: Kötü	0.0071	-0.0016	0.0012	-0.0009	0.0009
Yaş	0.0043	0.0000	-0.0001	0.0000	-0.0001

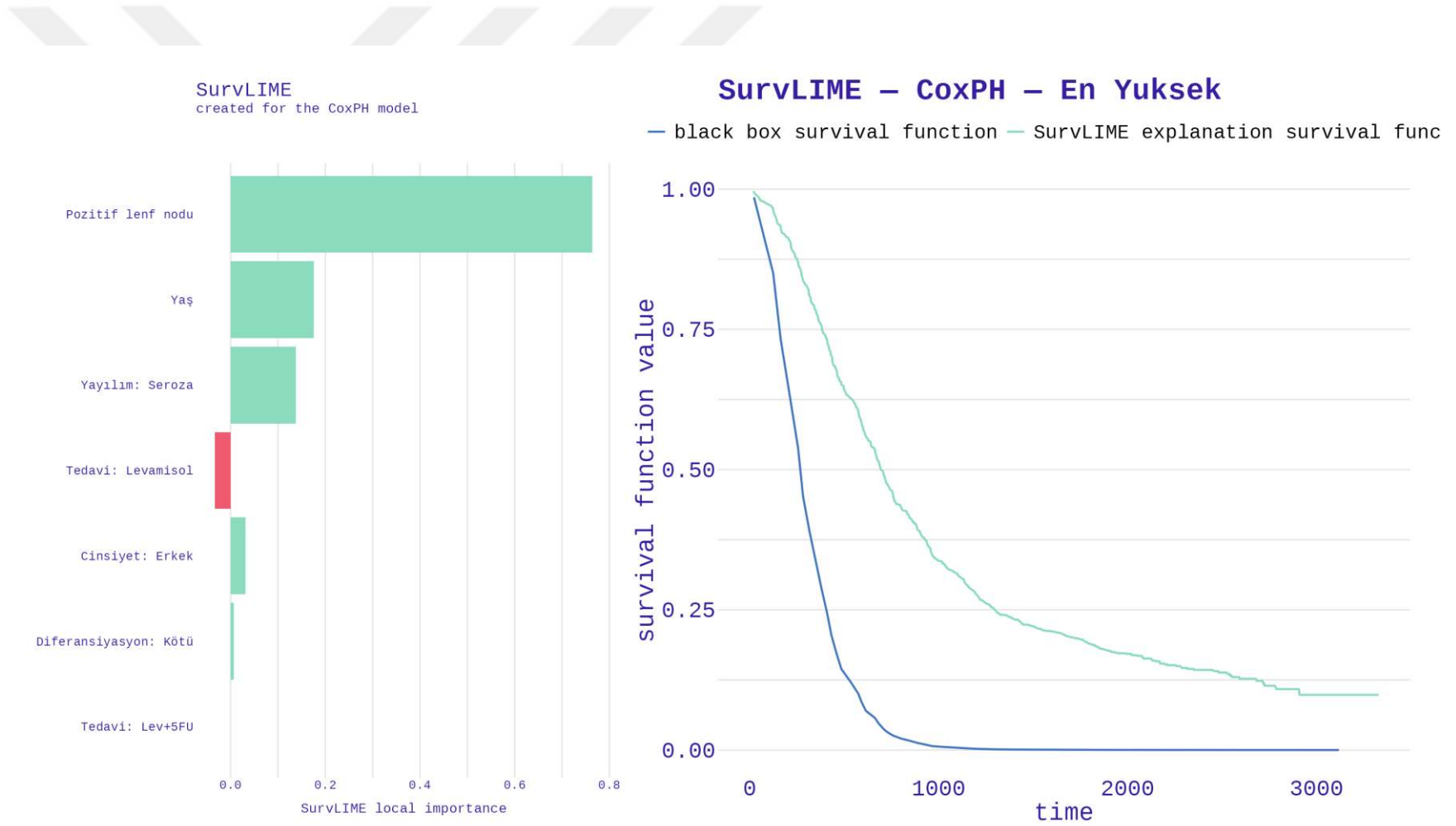
Tablo 4.7'de RSF modeline ait SurvLIME katsayıları yer aldı. CoxPH ile kıyaslandığında, RSF katsayılarının hem çok daha küçük ölçekli (mutlak değerler genellikle 0.02'nin altında) hem de risk düzeyleri arasında daha heterojen bir dağılım sergilediği dikkat çekti. Mutlak ortalama katkısı en yüksek öznitelik Perforasyon: Evet olup, özellikle orta (0.0228), düşük (0.0215) ve en düşük (0.0142) riskli gözlemlerde belirgin pozitif katsayılar aldı; bunu Obstrüksiyon: Evet ve Diferansiyasyon: Orta izledi. Tek bir özneliğin tüm risk düzeylerinde tutarlı biçimde baskın olmaması, ensemble modelin doğrusal olmayan karar mekanizmasının SurvLIME'in doğrusal yaklaşıklama varsayımıyla yalnızca kısmen temsil edilebildiğini ve yöntemin ağaç tabanlı modellerdeki sınırlılığını yansıttı.

Tablo 4.7. SurvLIME katsayıları: RSF modeli (4 gözlem)

Öznitelik	En Yüksek	Orta	Düşük	En Düşük
Perforasyon: Evet	0.0015	0.0228	0.0215	0.0142
Obstrüksiyon: Evet	0.0003	0.0067	0.0103	0.0129
Diferansiyasyon: Orta	-0.0022	-0.0069	-0.0078	-0.0095
Cerrahi süresi: Uzun	-0.0006	0.0090	0.0074	0.0005
Yayılım: Kas tabakası	-0.0014	-0.0086	0.0022	-0.0052
Diferansiyasyon: Kötü	0.0004	0.0017	0.0042	0.0105
Tedavi: Lev+5FU	0.0009	-0.0077	0.0044	-0.0036
Yayılım: Bitişik yapılar	-0.0018	-0.0022	0.0031	0.0071
Adezyon: Evet	0.0004	0.0063	0.0037	0.0015
Cinsiyet: Erkek	-0.0003	-0.0002	-0.0057	-0.0035
Yayılım: Seroza	-0.0012	-0.0040	-0.0025	0.0001
Tedavi: Levamisol	-0.0005	0.0017	0.0002	0.0025
Pozitif lenf nodu	-0.0000	0.0002	0.0002	0.0002
Yaş	0.0000	-0.0002	-0.0000	-0.0002

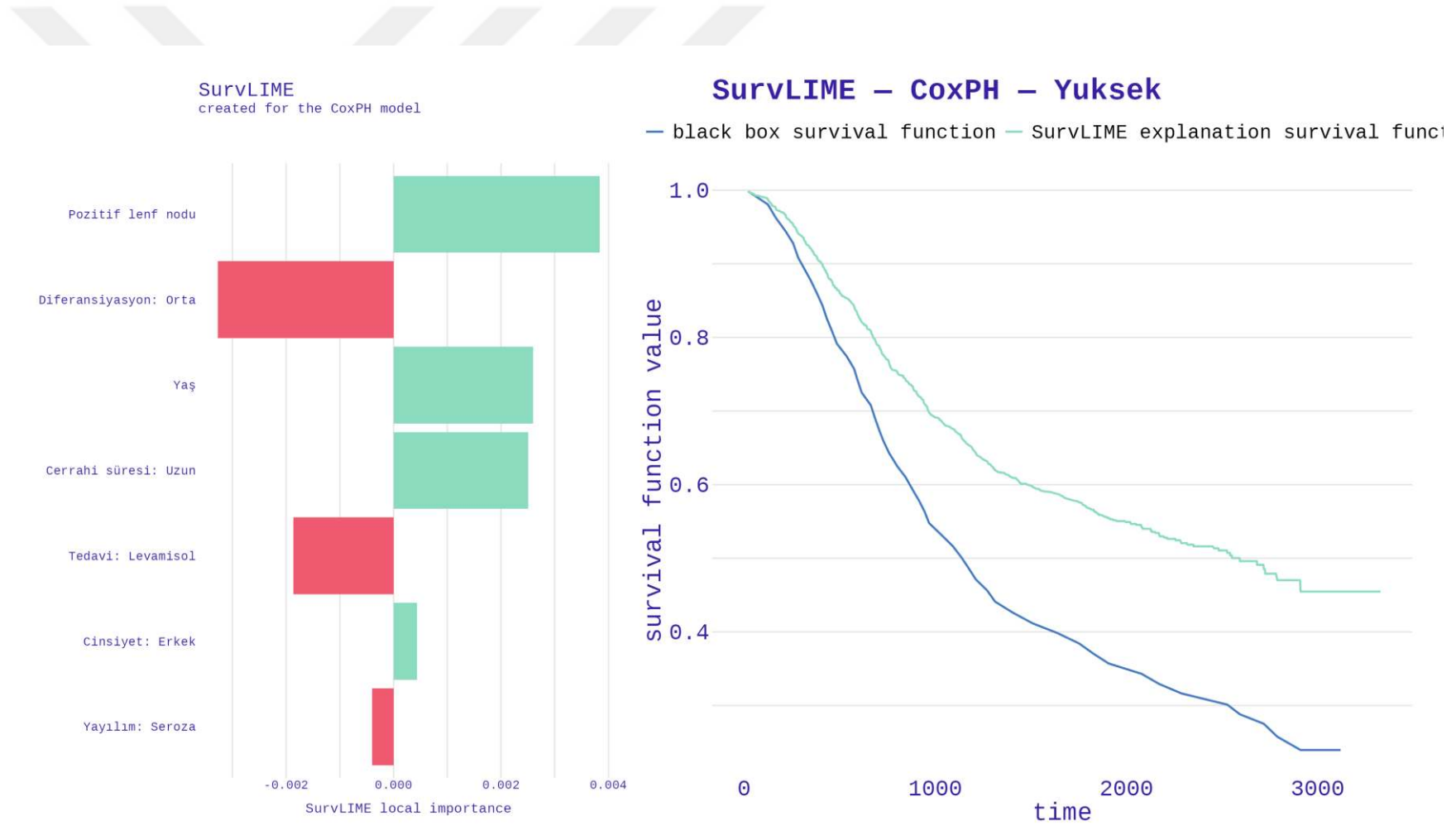
Şekil 4.57'de en yüksek riskli hasta için CoxPH modelinin SurvLIME açıklaması sunuldu. Sol panelde survex'in yerel önem (local importance) ölçütü yer almakta; bu uç risk profilinde sürekli Pozitif lenf nodu ve Yaş değişkenleri en büyük yerel katkıyı verdi (yerel Cox katsayıları için bk. Tablo 4.6). Sağ panelde SurvLIME açıklama eğrisi kara kutu modelin bir miktar üzerinde seyretmekte, bu da doğrusal yerel yaklaşıklamanın bu uç profilde sağkalımı kısmen olduğundan yüksek tahmin ettiğini gösterdi.





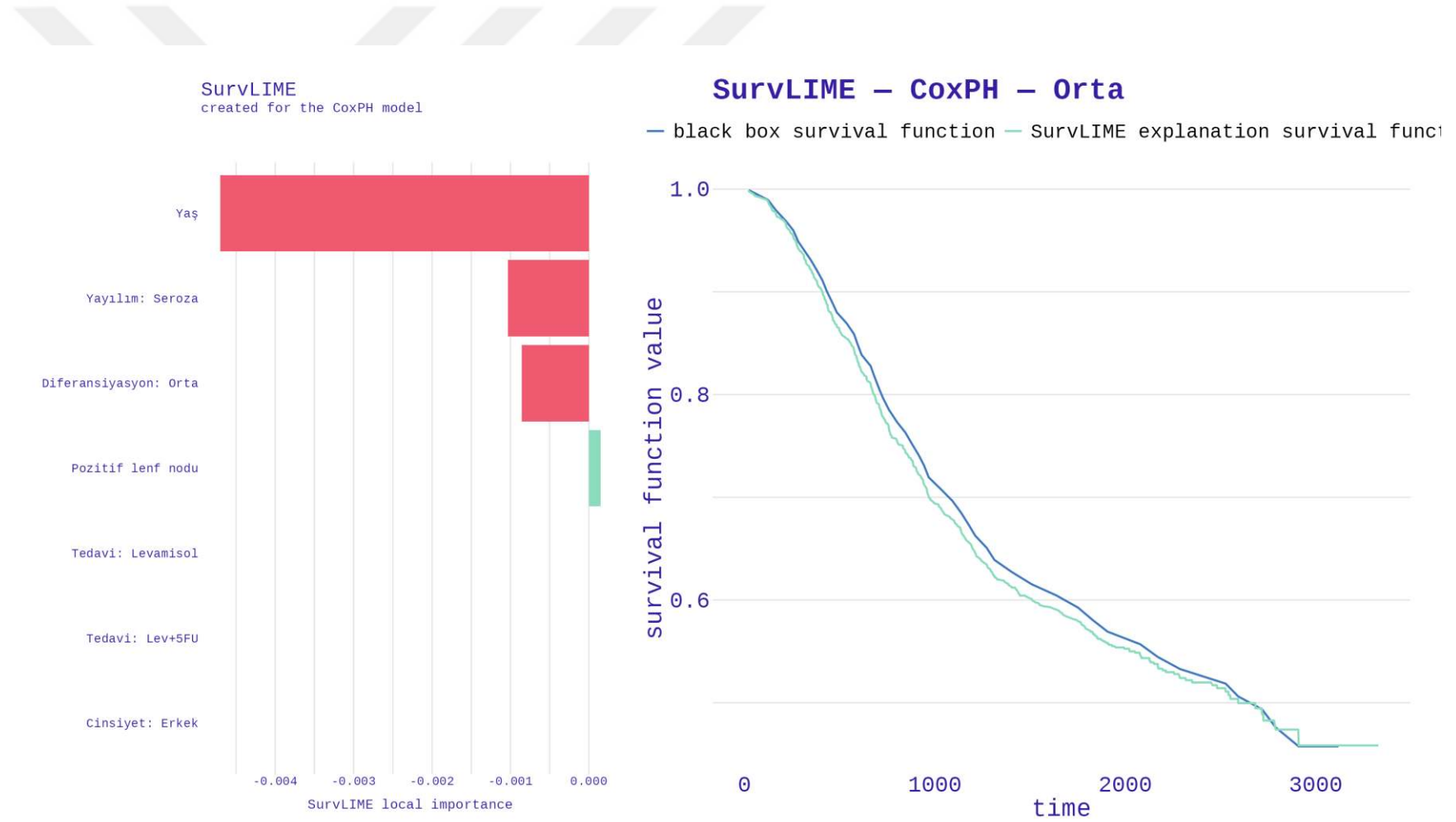
Şekil 4.57. CoxPH modeline ait SurvLIME en yüksek riskli hasta profili

Şekil 4.58'de yüksek riskli hasta profilinde yerel önem büyüklüklerinin en yüksek riskli profile kıyasla belirgin biçimde azaldığı; SurvLIME açıklama eğrisinin ise kara kutu eğriye yaklaştığı gözlemlendi.



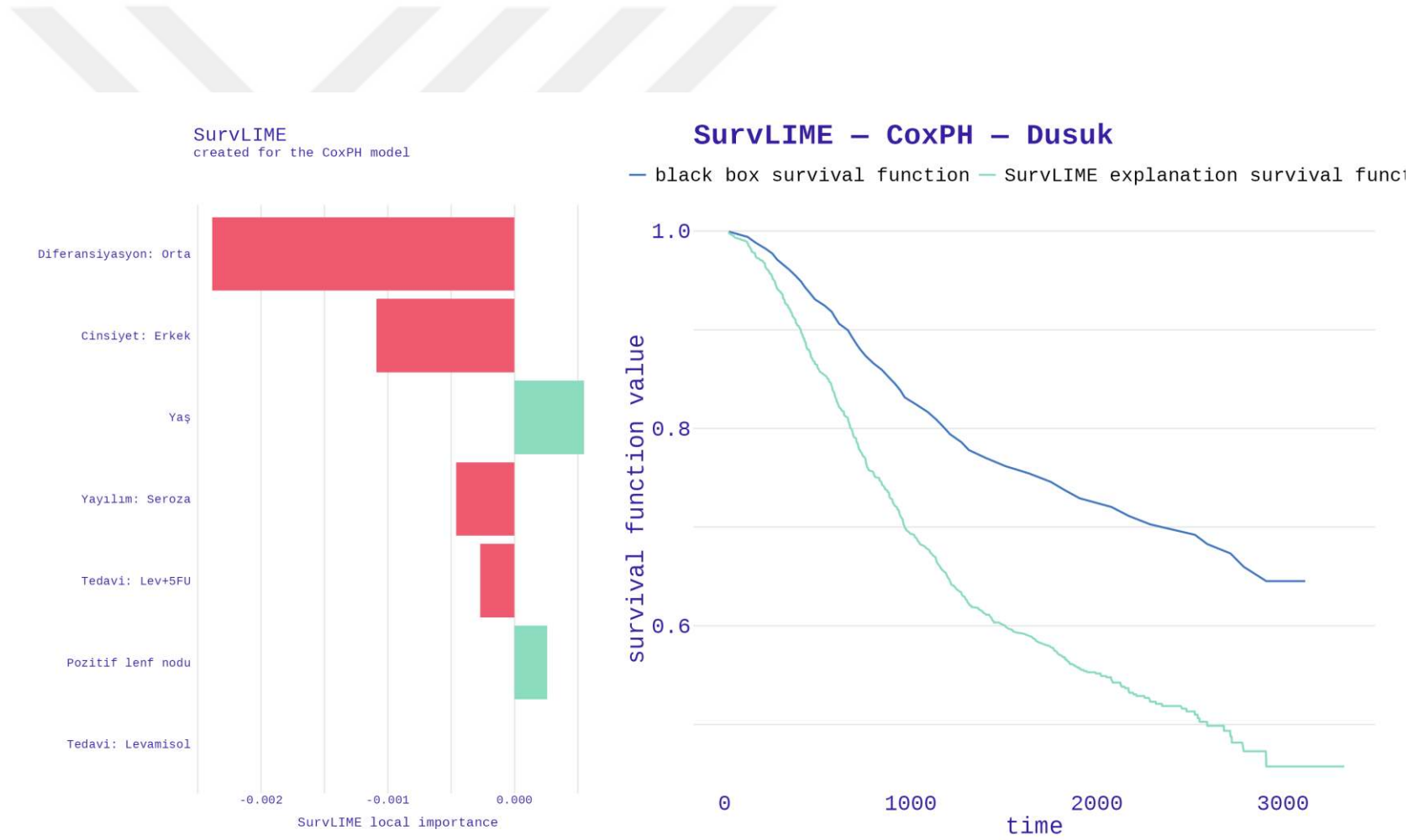
Şekil 4.58. CoxPH modeline ait SurvLIME yüksek riskli hasta profili

Şekil 4.59'da orta risk düzeyindeki hastada yerel önem değerlerinin düşük ve görece dengeli dağıldığı, herhangi bir öz niteliğin tek başına baskın etki göstermediği ve iki sağkalım eğrisinin yakın seyrettiği görüldü.



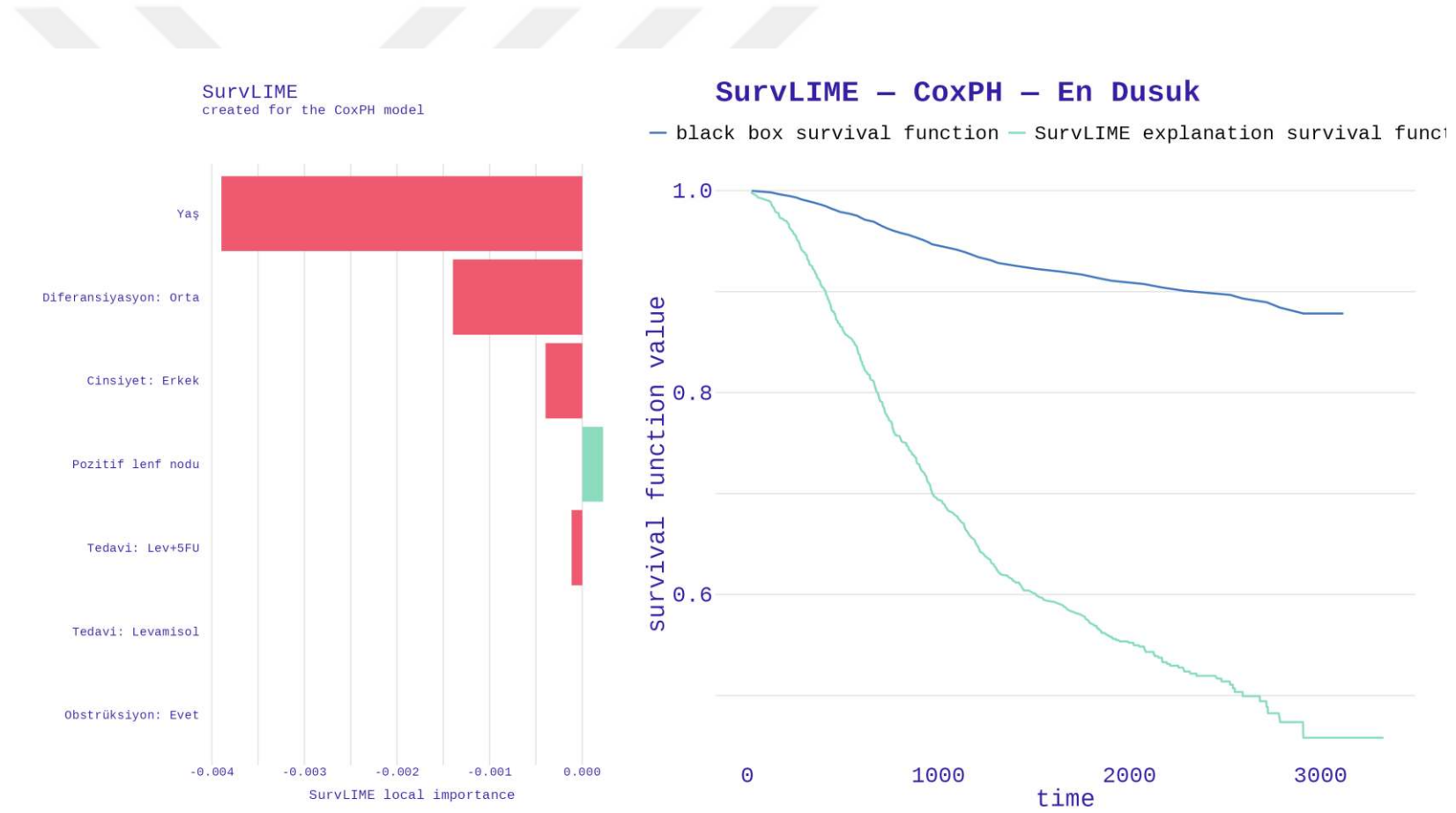
Şekil 4.59. CoxPH modeline ait SurvLIME orta riskli hasta profili

Şekil 4.60'ta düşük riskli hasta profilinde yerel önem büyüklükleri küçülmekte; SurvLIME açıklama eğrisi ile kara kutu eğri arasındaki uyumun risk azaldıkça değişkenlik gösterdiği izlendi.



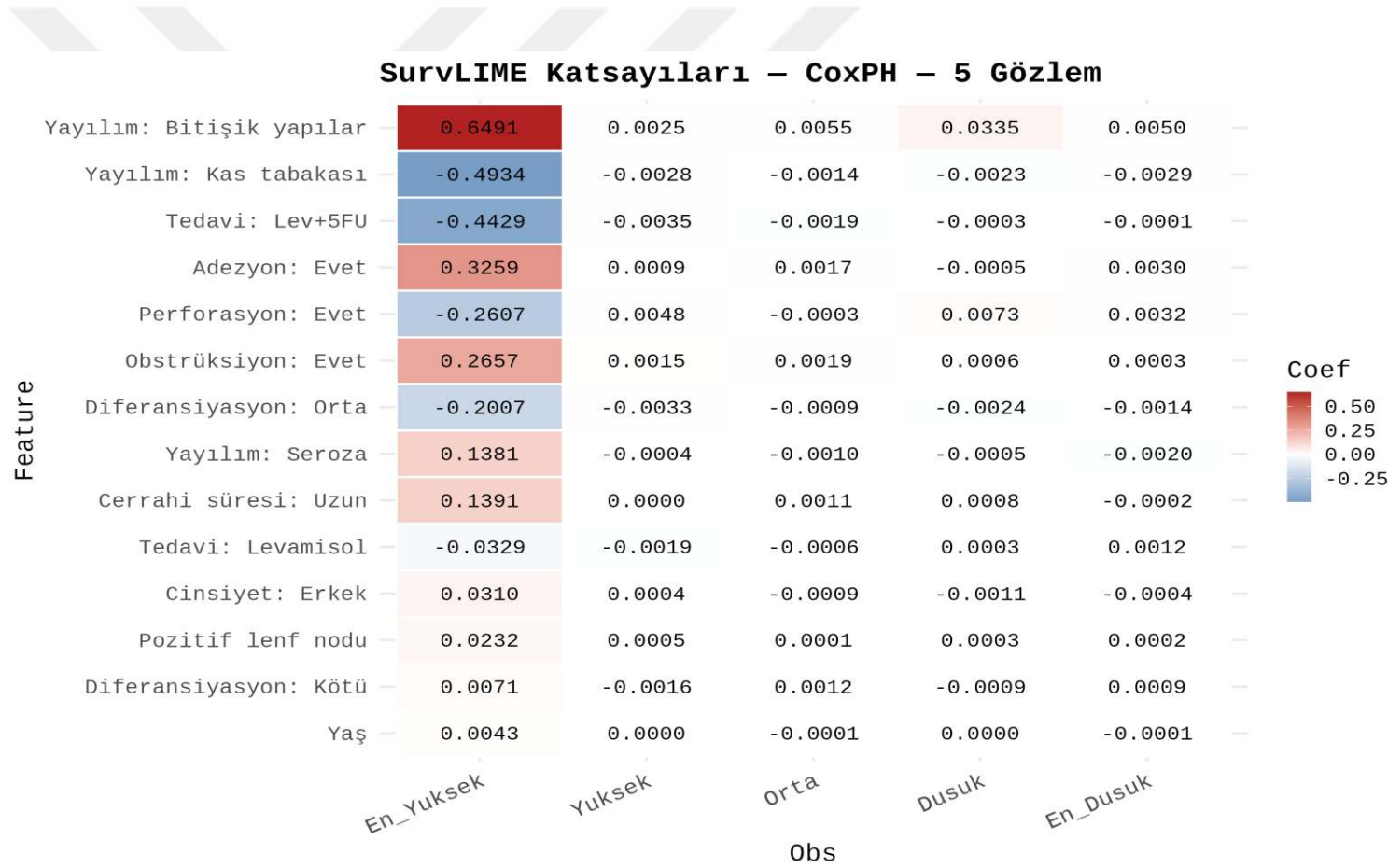
Şekil 4.60. CoxPH modeline ait SurvLIME düşük riskli hasta profili

Şekil 4.61'de en düşük riskli hastada yerel Cox katsayıları sıfıra çok yaklaşımakta (Tablo 4.6); buna karşılık SurvLIME açıklama eğrisi kara kutu eğriden aşağı yönde sapmakta, yani doğrusal yaklaşıklama bu uç profilde sağkalımı olduğundan düşük tahmin etti.



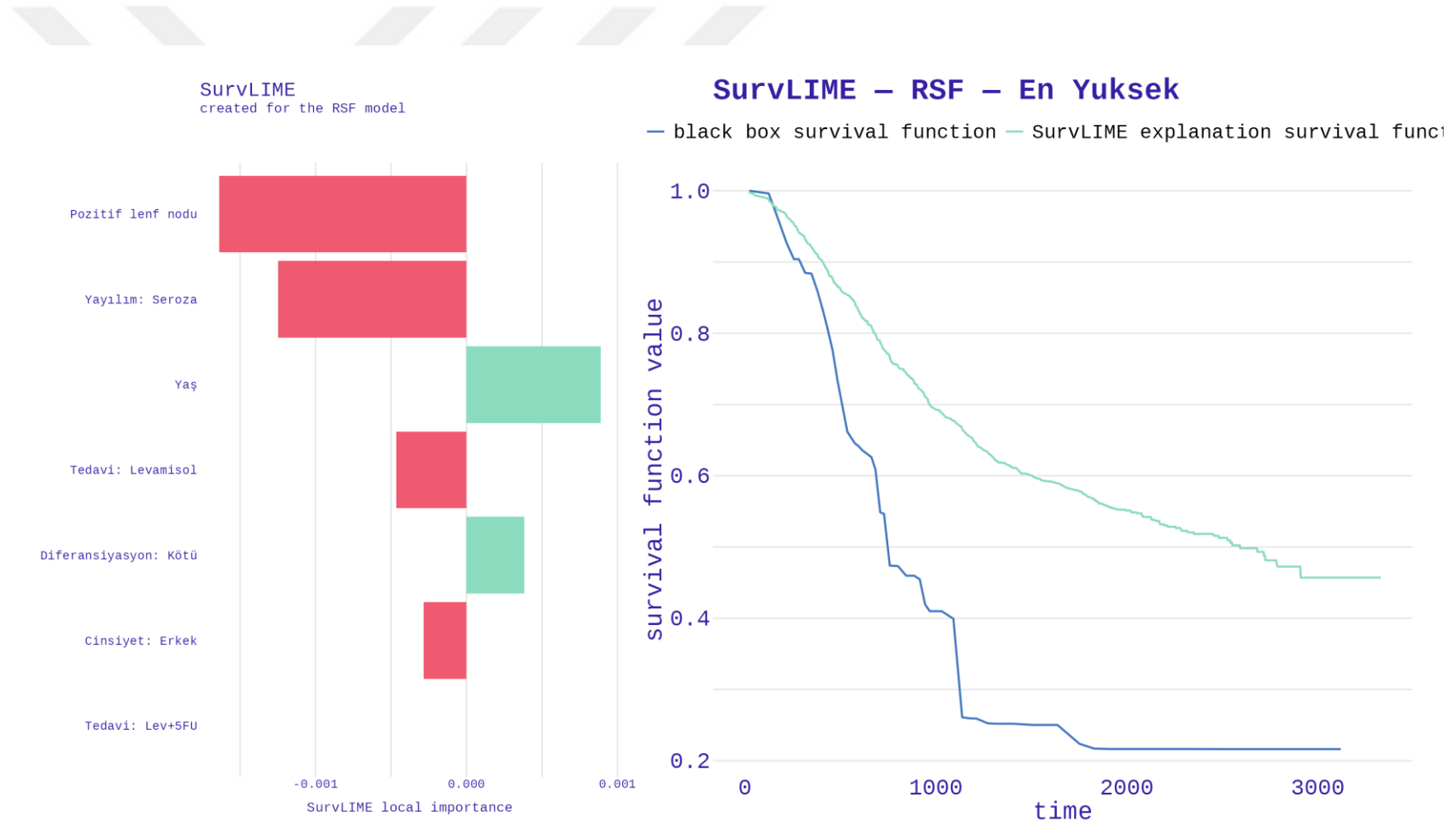
Şekil 4.61. CoxPH modeline ait SurvLIME en düşük riskli hasta profili

Şekil 4.62'de CoxPH modeline ait beş gözlemin SurvLIME katsayıları (β) karşılaştırmalı olarak sunuldu. Isı haritası, en yüksek riskli gözlemde katsayıların mutlak büyüklüğünün diğer risk düzeylerine kıyasla belirgin biçimde yüksek olduğunu; risk azaldıkça katsayıların sıfıra yaklaştığını ortaya koydu.



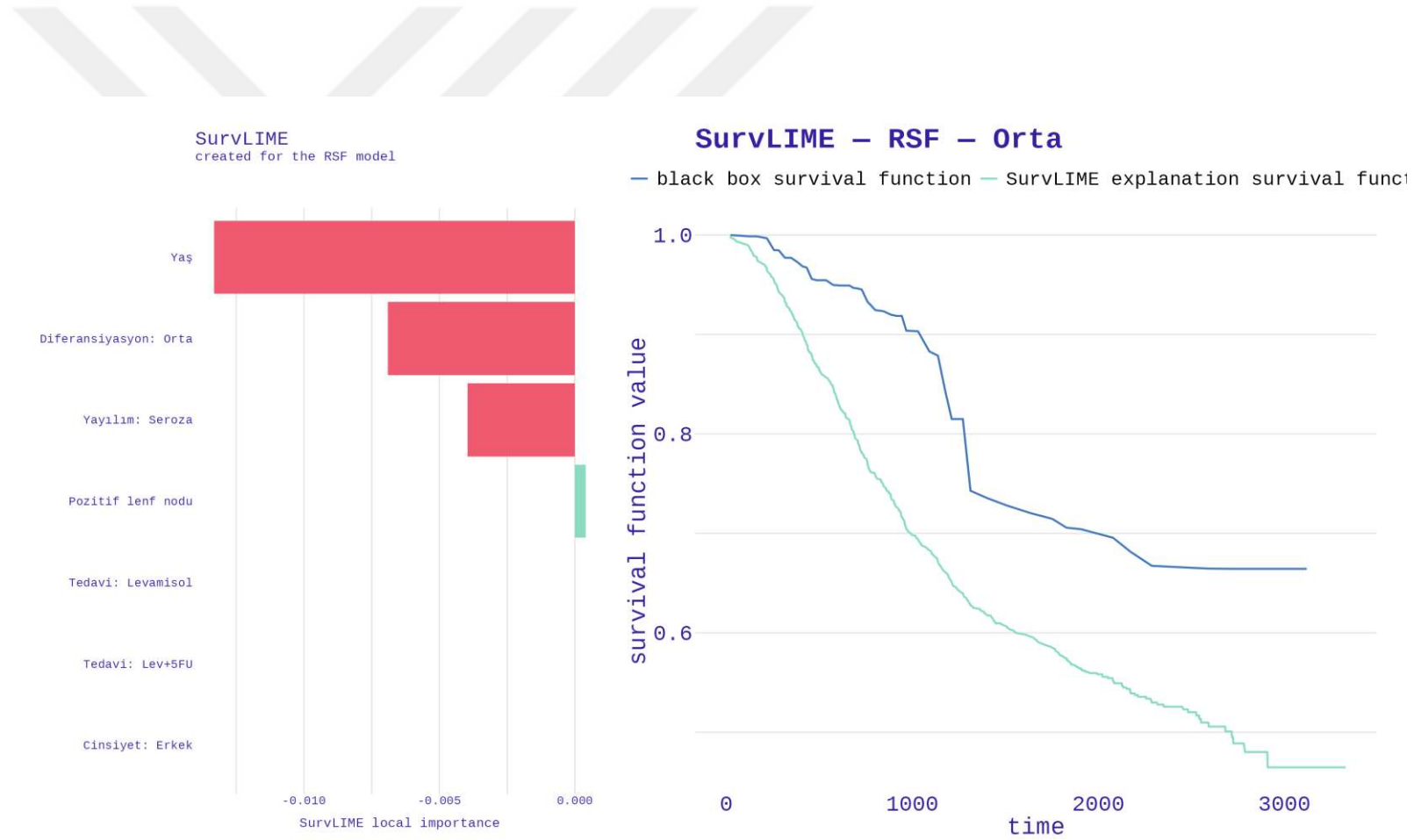
Şekil 4.62. CoxPH modeline ait SurvLIME katsayıları ısı haritası (5 gözlem)

Şekil 4.63'te RSF modelinin en yüksek riskli hasta için SurvLIME açıklaması yer aldı. RSF'nin yerel önem değerleri CoxPH'ye kıyasla çok daha küçük ölçeklidir; sağ panelde açıklama eğrisi kara kutu eğrinin üzerinde seyrederek doğrusal yaklaşıklamanın ağaç tabanlı modelin keskin risk düşüşlerini tam izleyemediğini gösterdi (katsayılar için bk. Tablo 4.7).



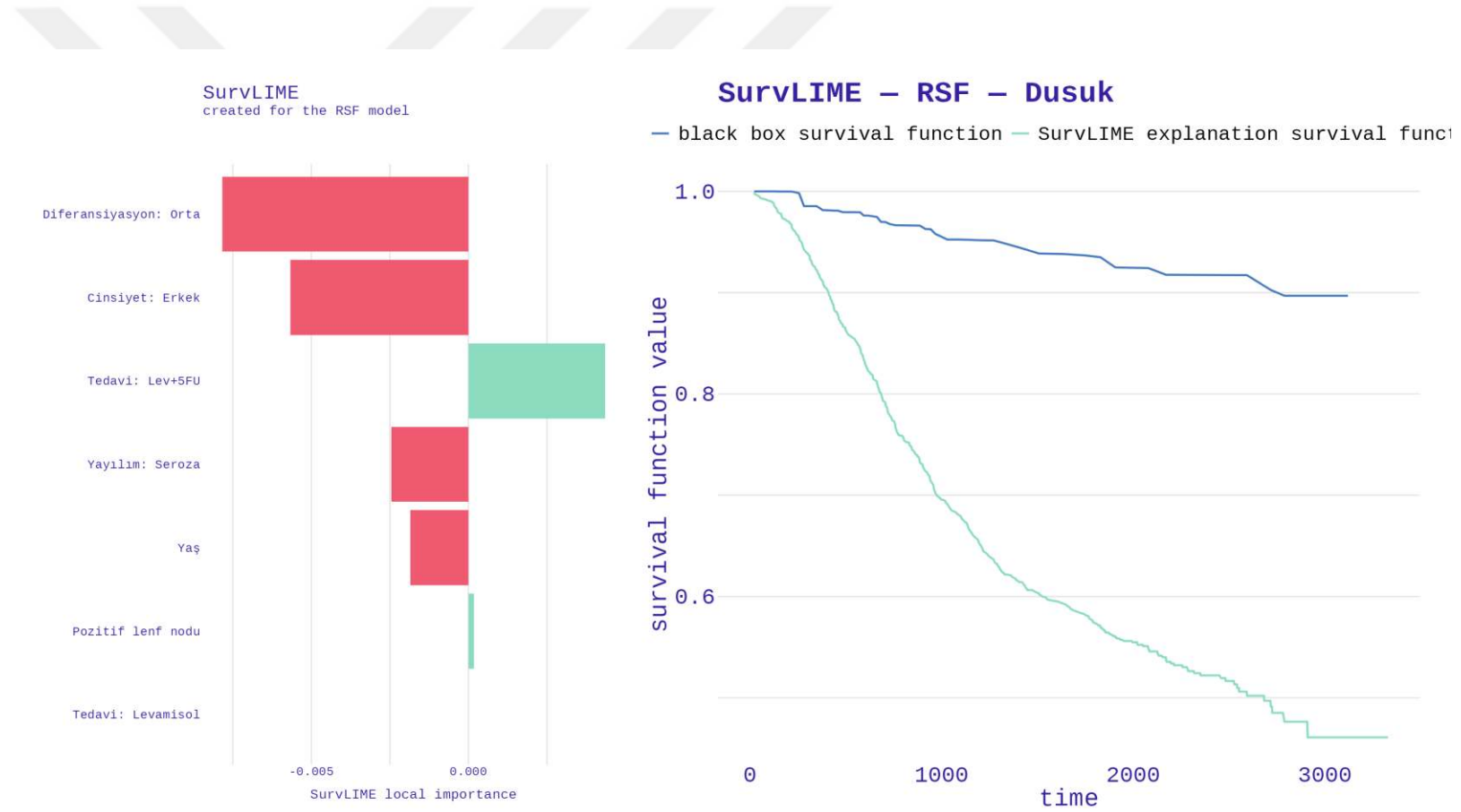
Şekil 4.63. RSF modeline ait SurvLIME en yüksek riskli hasta profili

Şekil 4.64'te orta riskli hastada RSF için Perforasyon: Evet ve Obstrüksiyon: Evet özniteliklerinin görece belirgin pozitif katsayılar aldığı, iki sağkalım eğrisinin ise yakın seyrettiği gözlemlendi.



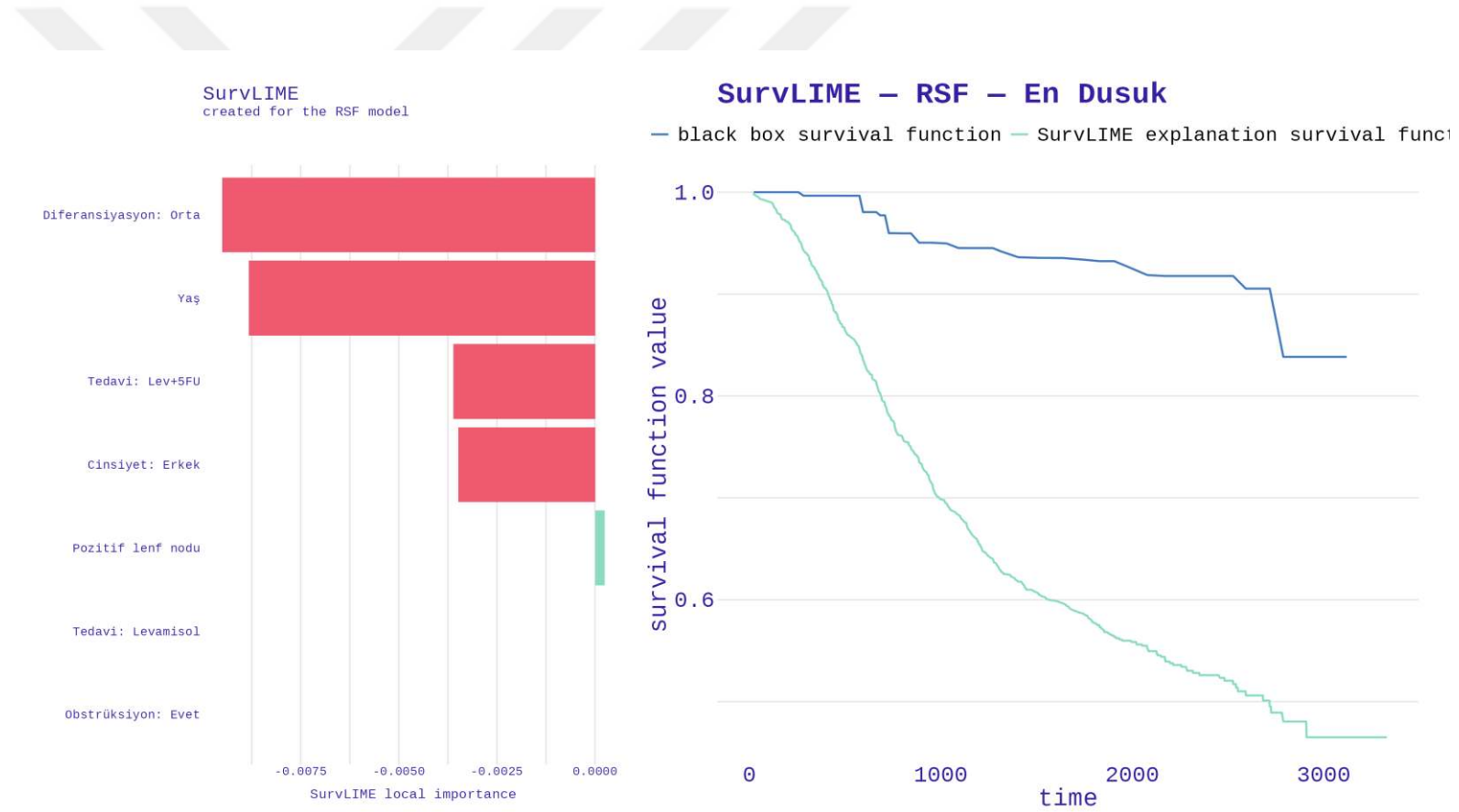
Şekil 4.64. RSF modeline ait SurvLIME orta riskli hasta profili

Şekil 4.65'te düşük riskli hastada RSF katsayılarının küçük ve heterojen dağıldığı; Perforasyon: Evet ve Obstrüksiyon: Evet değişkenlerinin pozitif yönde öne çıktığı görüldü.



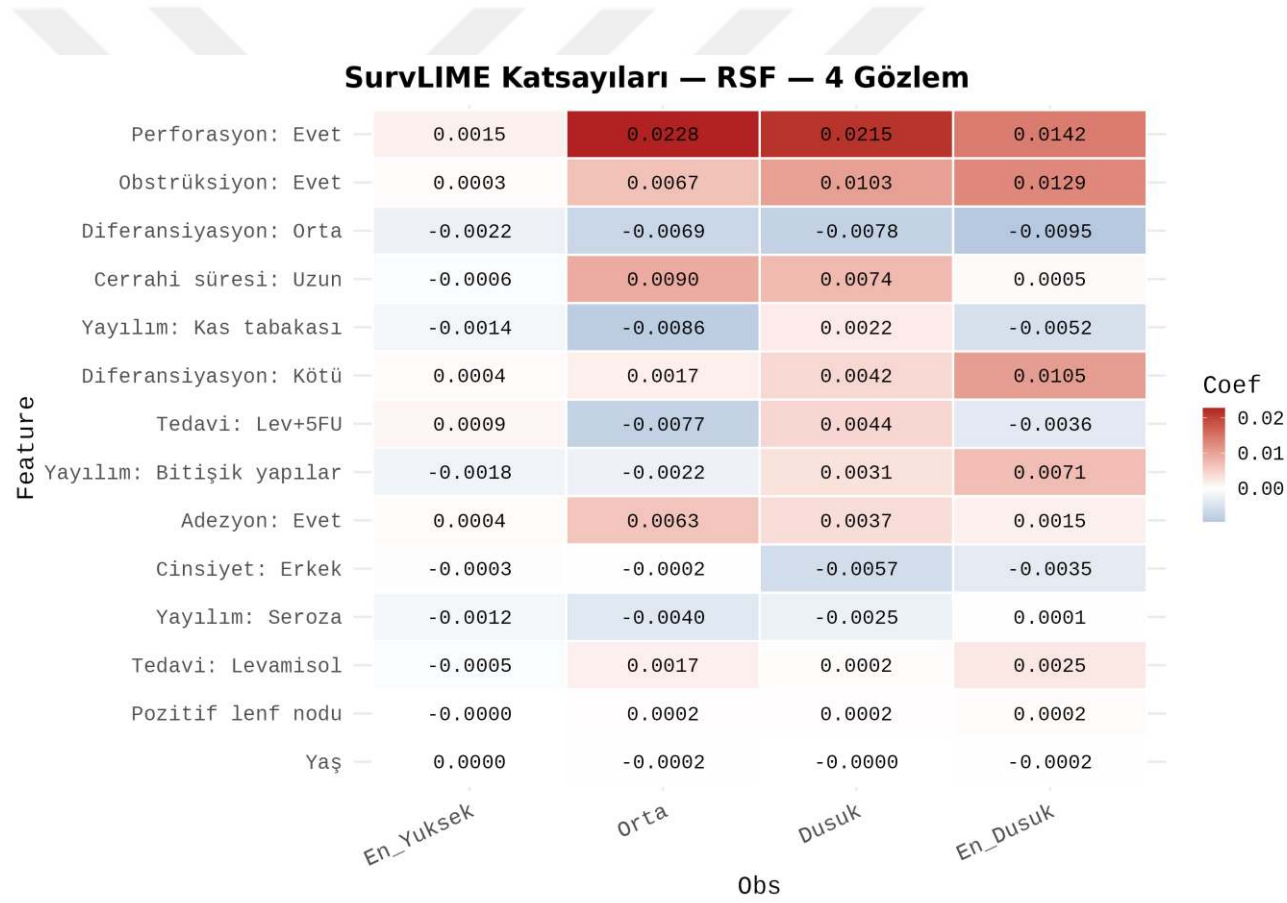
Şekil 4.65. RSF modeline ait SurvLIME düşük riskli hasta profili

Şekil 4.66'da en düşük riskli hastada Perforasyon: Evet ve Obstrüksiyon: Evet değişkenlerinin görece yüksek pozitif katsayılar sergilediği görüldü. Bu örüntü, SurvLIME'in doğrusal yaklaşıklama çerçevesinin ensemble modellerin karmaşık karar sınırlarını sınırlı biçimde temsil edebildiğini yansıttı.



Şekil 4.66. RSF modeline ait SurvLIME en düşük riskli hasta profili

Şekil 4.67'de RSF modeline ait dört gözlemin SurvLIME katsayıları görselleştirildi. CoxPH ısı haritasıyla karşılaştırıldığında, RSF'nin hem çok daha küçük ölçekli hem de risk düzeyleri arasında daha heterojen bir katsayı örüntüsü sergilediği, tek bir özneliğin tutarlı biçimde baskın olmadığı görüldü. Bu heterojenlik, ağaç tabanlı modelin doğrusal olmayan yapısının doğrusal yerel yaklaşıklama ile tam temsil edilememesinden kaynaklandı.



Şekil 4.67. RSF modeline ait SurvLIME katsayıları ısı haritası (4 gözlem)

5. TARTIŞMA

Bu çalışmada kolon kanseri sağkalım tahmininde on farklı istatistiksel, makine öğrenmesi ve derin öğrenme modeli karşılaştırıldı; üç farklı açıklanabilirlik yöntemi (SHAP, SurvSHAP(t), SurvLIME) ile model karar mekanizmaları incelendi. Elde edilen bulgular performans, açıklanabilirlik ve sınırlılıklar olmak üzere üç başlık altında tartışıldı.

Nested CV sonuçlarında modeller arası C-index farkı oldukça küçük kaldı ($\Delta C = 0.021$; en yüksek DeepSurv 0.6569, en düşük CART 0.6363). Harrell'in C-indexinin 0.66 düzeyinde olması kolon kanseri sağkalım tahmininin orta zorlukta bir problem olduğuna işaret etmektedir (58). Christodoulou ve arkadaşları (53) klinik tahmin problemlerinde karmaşık modellerin basit alternatiflere belirgin üstünlük sağlayamadığını göstermiştir; bu çalışmada da DeepSurv, RSF, DeepHit ve GradientBoosting modellerinin C-index değerleri 0.655-0.657 aralığında yakın konumlanmıştır.

Bootstrap analizinde sıralama değişti; ExtraSurvTrees (0.7350) ve DeepSurv (0.7278) öne çıktı. ExtraSurvTrees'in nested CV'de 9. sırada iken bootstrap'ta birinci olması, tüm veriye aşırı uyum (overfitting) kaynaklı iyimser yanlılığa işaret etmekte ve birincil değerlendirmenin nested CV üzerinden yapılması gerekliliğini güçlendirmektedir. Varma ve Simon'un (22) belirttiği gibi nested CV'nin eğitim-test ayrımı varyansı yansıtırken bootstrap tüm veri üzerindeki performansı ölçer; bu yöntemsel fark sıralama değişimini açıklamaktadır. ExtraSurvTrees'in en dar güven aralığına (0.0436) sahip olması ve KNN'in en geniş aralığa (0.0512) sahip olması, ensemble yöntemlerin tutarlı performansını, mesafe tabanlı yöntemlerin ise tahmin varyansını artırdığını göstermektedir.

CART modelinin en düşük C-index (0.6363) ile en zayıf ayırım gücünü göstermesi tek ağaç yapısının bilinen yüksek varyans sorununu yansıtmaktadır (62). Bu durum bagging tabanlı ensemble yöntemlerin (RSF) temel motivasyonunu doğrulamaktadır. IBS değerleri incelendiğinde doğrudan sağkalım fonksiyonu üreten beş model arasında RSF en düşük IBS'ye (0.1743) sahiptir. Breslow wrapper ile hesaplanan modellerde (DeepSurv, SurvivalSVM, KNN, DeepHit, XGBoost) IBS değerlerinin daha yüksek çıkması, Breslow baseline yaklaşımının orantılı tehlike varsayımına dayanmasının getirdiği bir sınırlılıktır (47).

DeepSurv'ün fold bazında belirgin değişkenlik göstermesi (C-index birinci ve ikinci foldlarda ~0.69'a ulaşırken diğer foldlarda ~0.64'e gerilemektedir), van der Ploeg ve arkadaşlarının belirttiği gibi modern modelleme tekniklerinin yeterli veriye ihtiyaç duyduğu ve n=888 gözlemin derin öğrenme modelleri için sınırdaki olduğu gözlemini somutlaştırmaktadır (49). Her iki derin öğrenme modelinin (DeepSurv, DeepHit) nested CV'de rekabetçi ortalama sergilemesine rağmen fold bazlı kararsızlık göstermesi, bu ölçekte ensemble yöntemlerin daha güvenilir bir seçim olduğunu düşündürmektedir.

TD-AUC analizinde bootstrap değerlendirmesinde ExtraSurvTrees (0.8098) ve DeepSurv (0.7901), nested CV'de ise RSF (0.7161) ve SurvivalSVM (0.7162) en yüksek değerlere ulaşmıştır. Tüm modellerde AUC'nin ilk 600 günde 0.75-0.82, 1000 günden sonra 0.70-0.77 aralığına düşmesi, sansürleme oranının artmasıyla tahmin belirsizliğinin yükselmesini yansıtmaktadır (96). SurvivalSVM'in 1200 günden sonra en belirgin düşüşü sergilemesi, sıralama tabanlı kayıp fonksiyonunun uzun vadeli tahminlerdeki sınırlılığını göstermektedir.

SHAP analizinde tüm on modelde pozitif lenf düğümü sayısı (nodes) en önemli prognostik faktör olarak belirlendi. Bu bulgu Compton ve arkadaşlarının kolorektal kanser prognostik faktörleri derlemesi ve Moertel çalışması (30) (108) ile tutarlıdır. Modeller arası tutarlı sıralama, elde edilen bulgunun yöntemsel bir artefakt olmadığını gerçek bir biyolojik sinyali yansıttığını desteklemektedir. Kalın bağırsak perforasyonunun (perfor) tüm modellerde en düşük öneme sahip olması, düşük prevalansından (%3.0) kaynaklanmaktadır (44).

Model ailesine bağlı farklılıklar dikkat çekicidir. Pozitif lenf düğümü sayısı (nodes) tüm model ailelerinde tek başına en baskın değişken konumundadır; SHAP katkısı ağaç tabanlı modellerde en yüksek düzeye ulaşmış (CART %47.5, DeepHit %47.4, RSF %47.4), gradyan güçlendirme modellerinde de (GradientBoosting %42.6, XGBoost %42.6) ilk sırada yer almıştır (55, 66). Tedavi etkisinde Lev+5FU grubunun (Gözlem referans alınmıştır) CoxPH'de %17.4, KNN'de %16.9 iken ağaç tabanlı modellerde daha düşük kalması, doğrusal modellerin tedavi grupları arasındaki farkı ana etki olarak yakalarken, ağaç tabanlı modellerin bu etkiyi değişken etkileşimleri içinde dağıttığını göstermektedir (46).

SurvSHAP(t) agregasyon değerleri statik SHAP sıralamasını zaman boyutunda destekledi; her iki analizde de pozitif lenf düğümü sayısı (nodes) en yüksek değere sahiptir. Global SurvSHAP(t) grafiklerinde pozitif lenf düğümü sayısı etkisinin ilk 1000 günde hızla arttığı, sonrasında platoya ulaştığı gözlemlendi. Bu patern Şekil 3.2'deki KM

eğrisindeki ilk 500 gündeki dik düşüşle uyumlu olup erken dönem mortalitede lenf düğümü tutulumunun kritik rolünü yansıtmaktadır (108).

CoxPH'nin SurvSHAP(t) grafiğinde değişken sıralamasının zaman boyunca sabit kalması orantılı tehlike varsayımının doğal sonucudur (6). Ağaç tabanlı modellerin grafiklerinde zamana bağlı değişkenlik gözlenmesi bu modellerin orantılı olmayan tehlike ilişkilerini yakalayabildiğini göstermektedir. DeepHit'in basamaklı SurvSHAP(t) paterni ayırık zaman yapısının bir yansımasıdır (12). Bireysel SurvSHAP(t) profillerinde yüksek riskli hastalarda nodes değişkeninin pozitif, düşük riskli hastalarda negatif yönde etki göstermesi, lenf düğümü tutulumunun prognostik değerini hasta düzeyinde doğrulamaktadır (30).

SurvLIME analizinde en yüksek riskli hastada Yayılım: Bitişik yapılar değişkeninin en büyük pozitif yerel katsayıya (0.6491) sahip olması, bitişik yapılara invazyonun yerel Cox yaklaşımında sağkalım riskini artıran başat faktör olduğunu göstermektedir; Lev+5FU tedavisi (-0.4429) ve Yayılım: Kas tabakası (-0.4934) ise koruyucu yönde katkı vermektedir. Risk düzeyi azaldıkça katsayı büyüklüklerinin küçülmesi, SurvLIME yerel açıklamalarının hasta risk profiline duyarlı olduğunu ortaya koymaktadır. RSF modelinde katsayılar CoxPH'ye kıyasla çok daha küçük ve risk düzeyleri arasında daha heterojen olup mutlak ortalama katkısı en yüksek değişken Perforasyon: Evet'tir; bu durum ensemble modellerin doğrusal olmayan karar sınırlarının SurvLIME'in doğrusal yaklaşımına çerçevesiyle tam temsil edilememesinden kaynaklanmaktadır (19). Ayrıca survex'in bireysel SurvLIME şekillerinde sunduğu yerel önem (local importance) ölçütü sürekli değişkenleri (ör. pozitif lenf nodu, yaş) uç risk profillerinde öne çıkarırken, Tablo 4.6 ve 4.7'deki ham katsayılar (β) kategorik göstergeleri vurgulamaktadır.

Bu çalışmanın başlıca sınırlılıkları aşağıda ele alınmaktadır. Birincisi, veri seti 888 gözlemden oluşmakta olup derin öğrenme modellerinin tam potansiyelini ortaya koyabilecek büyüklükte değildir (49); nitekim DeepSurv'ün fold bazında yüksek değişkenlik (sağa çarpık C-index dağılımı) göstermesi bu sınırlamayı yansıtmaktadır. İkincisi, SurvSHAP(t) hesaplamasında RSF ve ExtraSurvTrees modelleri için hesaplama maliyeti nedeniyle sınırlı sayıda gözlem ($n=2$) kullanılmıştır. Üçüncüsü, SurvLIME yalnızca R survex paketi (91) ile uygulanabilmiş olup Python ortamında solver uyumsuzluğu nedeniyle Python tabanlı modellerde çalıştırılamamıştır. Dördüncüsü, sağkalım fonksiyonu üretmeyen modellerde Breslow baseline tahmini kullanılmış olup bu yaklaşım orantılı tehlike varsayımına dayanmaktadır (47). Beşincisi, Tablo 4.5'teki

hiperparametre sonuçlarında CoxPH'nin yüksek regularizasyon ($\alpha=1.0$), GradientBoosting ve XGBoost'un düşük ağaç derinliği ($\text{max_depth}=2$), DeepSurv'ün küçük mimari (2 katman, 32 düğüm) tercih etmesi, 12 değişkenli ve $n=888$ ölçekli bu veri setinde aşırı uyumu engellemek için güçlü düzenleştirme gerekliliğini yansıtmaktadır (45, 55); bu durum bulguların farklı ölçekteki veri setlerine doğrudan genelleştirilebilirliğini sınırlamaktadır.



6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, R survival paketindeki colon veri seti ($n=888$, olay=430, olay oranı=%48.4) üzerinde on farklı makine öğrenmesi ve derin öğrenme modeli üç performans metriğiyle karşılaştırılmış ve üç açıklanabilirlik yöntemiyle değerlendirilmiştir.

İç içe çapraz doğrulama (5×3) sonuçlarına göre en yüksek ortalama C-index değerleri DeepSurv (0.6569 ± 0.0211), RSF (0.6568 ± 0.0188) ve DeepHit (0.6567 ± 0.0222) modellerinde birbirine çok yakın elde edilirken, en düşük değer CART modelinde (0.6363 ± 0.0227) gözlemlendi. Bootstrap analizi ($B=1000$) sonuçlarına göre ExtraSurvTrees (C-index=0.7350, %95 GA [0.7141-0.7577]) ve DeepSurv (C-index=0.7278, %95 GA [0.7064-0.7512]) en yüksek değerleri sergiledi; ancak özellikle ExtraSurvTrees'in nested CV sıralamasıyla çelişmesi bootstrap'ın iyimser yanlılığını yansıtmaktadır. Kalibrasyon açısından RSF modeli en düşük IBS değeriyle (0.1743 ± 0.0099) en iyi performansı gösterdi. Zamana bağlı AUC analizinde nested CV'de SurvivalSVM (TD-AUC=0.7162) ve RSF (TD-AUC=0.7161), bootstrap analizinde ise ExtraSurvTrees (0.8098) ve DeepSurv (0.7901) en yüksek değerlere ulaştı. Modeller arasındaki performans farkının küçük olması (Δ C-index=0.021), kolon kanseri sağkalım tahmininin orta zorlukta bir problem olduğuna ve tek bir modelin mutlak üstünlük sağlayamadığına işaret etmektedir. DeepSurv nested CV'de en yüksek ortalama C-index değerini elde etmesine rağmen fold bazında yüksek değişkenlik (sağa çarpık dağılım) göstermiş, bu durum $n=888$ gözlemin derin öğrenme modelleri için sınırda bir örneklem büyüklüğü olmasıyla ilişkilendirilmiştir.

Açıklanabilirlik analizlerinde SHAP yöntemi tüm modellerde tutarlı olarak pozitif lenf düğümü sayısı (nodes) değişkenini en önemli prognostik faktör olarak belirlemiştir. SurvSHAP(t) analizi öznetelik öneminin zamana bağlı değişimini ortaya koymuş; lenf düğümü etkisi ilk 1000 günde hızla artmış, sonrasında platoya ulaşmıştır. SurvLIME analizi ise CoxPH (beş gözlem) ve RSF (dört gözlem) modelleri üzerinde hasta bazlı yerel açıklamalar sunmuştur.

Bu çalışmada geliştirilen tahmin çerçevesi, kolon kanseri hastalarının bireysel risk profillerinin belirlenmesine yönelik somut araçlar sunmaktadır. SurvLIME analizi ile farklı risk düzeylerindeki hasta profilleri için elde edilen yerel açıklamalar, klinisyenlerin tedavi kararlarını destekleyebilecek yorumlanabilir çıktılar üretmiştir. Örneğin yüksek

riskli hasta profilinde derin tümör invazyonunun (bitişik yapılara yayılım) ve pozitif lenf düğümü tutulumunun sağkalımı azaltıcı yönde öne çıktığı, düşük riskli profillerde ise bu etkilerin belirgin biçimde zayıfladığı gösterildi. SHAP öznitelik önem sıralaması, cerrahide lenf düğümü değerlendirmesinin prognostik önemini nicel olarak doğrulamaktadır.

SurvLIME ile elde edilen bu tür bireysel (yerel) açıklamalar, klinik uygulamada çok yönlü fayda sağlama potansiyeli taşımaktadır. Hastaya özgü risk skorları ve bu skorları yönlendiren öznitelikler birlikte sunulduğunda, adjuvan tedavi (cerrahi sonrası uygulanan destekleyici kemoterapi) yoğunluğunun bireyselleştirilmesine olanak tanınır; yüksek riskli olarak sınıflandırılan hastalarda daha yoğun izlem ve agresif tedavi stratejileri gündeme alınırken, düşük riskli hastalarda gereksiz tedavi yükünden kaçınılabilir. Modelin yalnızca bir risk tahmini üretmekle kalmayıp bu tahmini hangi klinik değişkenlere dayandırıldığını şeffaf biçimde ortaya koyması, klinisyenlerin modele duyduğu güveni artırarak kara kutu yaklaşımlarının benimsenmesinin önündeki en önemli engeli azaltmaktadır. Ayrıca yorumlanabilir çıktılar, hasta-hekim iletişimde somut bir dayanak oluşturarak hastaların kendi prognozlarını ve risk faktörlerini anlamalarını kolaylaştırmakta ve ortak karar alma sürecini desteklemektedir. Bireysel risk düzeyine göre izlem sıklığının ve tanısal kaynakların önceliklendirilmesi, sınırlı sağlık kaynaklarının daha etkin kullanılmasına da katkı sağlayabilir. Bütün bu yönleriyle açıklanabilir sağkalım modelleri, tümör (T), bölgesel lenf nodu (N) ve uzak metastaz (M) durumuna dayanan mevcut TNM evreleme sistemini tamamlayıcı bir karar destek aracı olarak klinik karar destek sistemlerine entegre edilebilir ve bireyselleştirilmiş tedavi planlamasında kullanılabilir.

Bu çalışmanın bulguları doğrultusunda çeşitli öneriler geliştirilmiştir. İlk olarak, klinik karar destek sistemlerinde iç içe çapraz doğrulamada en dengeli performansı gösteren ensemble modeli RSF'nin (C-index=0.6568, IBS=0.1743, TD-AUC=0.7161) açıklanabilirlik yöntemleriyle birlikte kullanılması önerilmektedir. İkinci olarak, derin öğrenme modellerinin performansının $n>5000$ düzeyinde çok merkezli veri setlerinde yeniden değerlendirilmesi gerekmektedir. Üçüncü olarak, SHAP, SurvSHAP(t) ve SurvLIME yöntemlerinin farklı kanser türlerinde ve farklı veri yapılarında karşılaştırmalı olarak uygulanması bulguların genellenebilirliğini artıracaktır. Dördüncü olarak, zamana bağlı açıklanabilirlik yöntemlerinin bireyselleştirilmiş tedavi planlamasına entegrasyonu araştırılmalıdır. Son olarak, gelecek çalışmalarda rekabet eden riskler, zaman bağımlı kovaryatlar ve eksik veri senaryolarının modele dahil edilmesi önerilmektedir.

KAYNAKLAR

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, ve ark. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-63.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, ve ark. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-49.
3. Türkiye Kanser İstatistikleri. T.C. Sağlık Bakanlığı Türkiye Halk Sağlığı Genel Müdürlüğü Kanser Dairesi Başkanlığı. <https://hsgm.saglik.gov.tr/tr/kanser-istatistikleri> Son Erişim Tarihi 15 Ocak 2025.
4. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. 2023;73(1):17-48.
5. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-81.
6. Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol*. 1972;34(2):187-202.
7. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina*. 2020;56(9):455.
8. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3):841-60.
9. Ridgeway G. The state of boosting. *Comput Sci Stat*. 1999;31:172-81.
10. Van Belle V, Pelckmans K, Van Huffel S, Suykens JAK. Support vector machines for survival analysis. *Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare*. 2007:1-8.
11. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol*. 2018;18(1):24.
12. Lee C, Zame W, Yoon J, Van der Schaar M. DeepHit: a deep learning approach to survival analysis with competing risks. *Proc AAAI Conf Artif Intell*. 2018;32(1):2314-21.

13. Wiegrefe S, Kopper P, Sonabend R, Bischl B, Bender A. Deep learning for survival analysis: a review. *Artif Intell Rev.* 2024;57:65.
14. Wang P, Li Y, Reddy CK. Machine learning for survival analysis: a survey. *ACM Comput Surv.* 2019;51(6):1-36.
15. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 2019;1(5):206-15.
16. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30:4765-74.
17. Ribeiro MT, Singh S, Guestrin C. Why should I trust you?: explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD.* 2016:1135-44.
18. Krzyżiński M, Spytek M, Baniecki H, Biecek P. SurvSHAP(t): time-dependent explanations of machine learning survival models. *Knowl Based Syst.* 2023;262:110234.
19. Kovalev MS, Utkin LV, Kasimov EM. SurvLIME: a method for explaining machine learning survival models. *Knowl Based Syst.* 2020;203:106164.
20. Biecek P, Burzykowski T. *Explanatory Model Analysis.* New York, Chapman and Hall/CRC, 2021:1-312.
21. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access.* 2018;6:52138-60.
22. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics.* 2006;7:91.
23. Cawley GC, Talbot NLC. On over-fitting in model selection. *J Mach Learn Res.* 2010;11:2079-107.
24. Haider H, Hoehn B, Davis S, Greiner R. Effective ways to build and evaluate individual survival distributions. *J Mach Learn Res.* 2020;21(85):1-63.
25. Sonabend R, Király FJ, Bender A, Bischl B, Lang M. mlr3proba: an R package for machine learning in survival analysis. *Bioinformatics.* 2021;37(17):2789-91.
26. Kleinbaum DG, Klein M. *Survival Analysis: A Self-Learning Text*, 3. Bask1. New York, Springer, 2012:1-700.
27. Collett D. *Modelling Survival Data in Medical Research*, 3. Bask1. Boca Raton, CRC Press, 2015:1-548.
28. Hosmer DW, Lemeshow S, May S. *Applied Survival Analysis*, 2. Bask1. Hoboken, Wiley, 2008:1-392.

29. Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data, 2. Baskı. New York, Springer, 2003:1-536.
30. Compton CC, Fielding LP, Burgart LJ, ve ark. Prognostic factors in colorectal cancer. Arch Pathol Lab Med. 2000;124(7):979-94.
31. Huang Y, Li J, Li M, Aparasu RR. Application of machine learning in predicting survival outcomes involving real-world data: a scoping review. BMC Med Res Methodol. 2023;23(1):268.
32. Leung KM, Elashoff RM, Afifi AA. Censoring issues in survival analysis. Annu Rev Public Health. 1997;18(1):83-104.
33. Lagakos SW. General right censoring and its impact on the analysis of survival data. Biometrics. 1979;35(1):139-56.
34. Hougaard P. Frailty models for survival data. Lifetime Data Anal. 1995;1(3):255-73.
35. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data, 2. Baskı. Hoboken, Wiley, 2002:1-462.
36. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ. 1998;317(7172):1572.
37. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep. 1966;50(3):163-70.
38. Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. Biometrika. 1982;69(3):553-66.
39. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis. Br J Cancer. 2003;89(3):431-36.
40. Borgan Ø. Three contributions to the Encyclopedia of Biostatistics: the Nelson-Aalen, Kaplan-Meier, and Aalen-Johansen. Department of Mathematics, University of Oslo. 1997:1-12.
41. Cox DR, Oakes D. Analysis of Survival Data. London, Chapman and Hall, 1984:1-201.
42. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York, Springer, 2000:1-350.
43. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515-26.
44. Steyerberg EW. Clinical Prediction Models, 2. Baskı. Cham, Springer, 2019:1-558.

45. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16(4):385-95.
46. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw*. 2011;39(5):1-13.
47. Breslow NE. Discussion of the paper by DR Cox. *J R Stat Soc Series B Stat Methodol*. 1972;34(2):216-17.
48. Wolpert DH, Macready WG. No free lunch theorems for optimization. *IEEE Trans Evol Comput*. 1997;1(1):67-82.
49. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry. *BMC Med Res Methodol*. 2014;14(1):137.
50. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*, 2. Baskı. New York, Springer, 2009:1-745.
51. Ishwaran H, Kogalur UB. Random survival forests for R. *R News*. 2007;7(2):25-31.
52. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8-17.
53. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22.
54. Dietterich TG. Ensemble methods in machine learning. *International Workshop on Multiple Classifier Systems*. 2000:1-15.
55. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189-232.
56. Kos FT, Cecen Kaynak S, Aktürk Esen S, Arslan H, Uncu D. Comparison of different machine learning models for predicting long-term overall survival in non-metastatic colorectal cancers. *Cureus*. 2024;16(12):e75713.
57. Breiman L, Friedman J, Olshen RA, Stone CJ. *Classification and Regression Trees*. Belmont, Wadsworth, 1984:1-358.
58. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543-46.
59. LeBlanc M, Crowley J. Survival trees by goodness of split. *J Am Stat Assoc*. 1993;88(422):457-67.

60. Bou-Hamad I, Larocque D, Ben-Ameur H. A review of survival trees. *Stat Surv.* 2011;5:44-71.
61. Hothorn T, Bühlmann P, Dudoit S, Molinaro A, Van Der Laan MJ. Survival ensembles. *Biostatistics.* 2006;7(3):355-73.
62. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5-32.
63. Ishwaran H, Kogalur UB, Chen X, Minn AJ. Random survival forests for high-dimensional data. *Stat Anal Data Min.* 2011;4(1):115-32.
64. Pölsterl S. scikit-survival: a library for time-to-event analysis built on top of scikit-learn. *J Mach Learn Res.* 2020;21(212):1-6.
65. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, ve ark. XGBoost: extreme gradient boosting. R Package Version 1.7.5.1. 2023.
66. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD.* 2016:785-94.
67. Barnwal A, Cho H, Hocking T. Survival regression with accelerated failure time model in XGBoost. *J Comput Graph Stat.* 2022;31(4):1292-302.
68. Cover TM, Hart PE. Nearest neighbor pattern classification. *IEEE Trans Inf Theory.* 1967;13(1):21-27.
69. Lowsky DJ, Ding Y, Lee DKK, McCulloch CE, Ross LF, Thistlethwaite JR, ve ark. A K-nearest neighbors survival probability prediction method. *Stat Med.* 2013;32(12):2062-69.
70. Beyer K, Goldstein J, Ramakrishnan R, Shaft U. When is nearest neighbor meaningful? *Proceedings of the 7th International Conference on Database Theory.* 1999:217-35.
71. Vapnik VN. *The Nature of Statistical Learning Theory*, 2. Baskı. New York, Springer, 2000:1-314.
72. Van Belle V, Pelckmans K, Van Huffel S, Suykens JAK. Improved performance on high-dimensional survival data by application of Survival-SVM. *Bioinformatics.* 2011;27(1):87-94.
73. Goodfellow I, Bengio Y, Courville A. *Deep Learning.* Cambridge, MIT Press, 2016:1-775.
74. Nagpal C, Li X, Dubrawski A. Deep survival machines: fully parametric survival regression and representation learning for censored data with competing risks. *IEEE J Biomed Health Inform.* 2021;25(8):3163-75.

75. Faraggi D, Simon R. A neural network model for survival data. *Stat Med*. 1995;14(1):73-82.
76. Kvamme H, Borgan Ø, Scheel I. Time-to-event prediction with neural networks and Cox regression. *J Mach Learn Res*. 2019;20(129):1-30.
77. Ching T, Zhu X, Garmire LX. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput Biol*. 2018;14(4):e1006076.
78. Lee C, Yoon J, Van der Schaar M. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks. *IEEE Trans Biomed Eng*. 2020;67(1):122-33.
79. Kvamme H, Borgan Ø. Continuous and discrete-time survival prediction with neural networks. *Lifetime Data Anal*. 2021;27(4):710-36.
80. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res*. 2012;13:281-305.
81. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. *IEEE Trans Neural Netw Learn Syst*. 2021;32(11):4793-813.
82. Goodman B, Flaxman S. European Union regulations on algorithmic decision-making and a right to explanation. *AI Magazine*. 2017;38(3):50-57.
83. Molnar C. *Interpretable Machine Learning*, 2. Baskı. 2022. <https://christophm.github.io/interpretable-ml-book> Son Erişim Tarihi 20 Ocak 2025.
84. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Comput Surv*. 2018;51(5):1-42.
85. Utkin LV, Satyukov ED, Konstantinov AV. SurvNAM: the machine learning survival model explanation. *Neural Netw*. 2022;147:81-102.
86. Lundberg SM, Erion GG, Lee SI. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*. 2018.
87. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. *Proceedings of the Machine Learning for Healthcare Conference*. 2019:359-80.
88. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care. *J Biomed Inform*. 2021;113:103655.

89. Shapley LS. A value for n-person games. İçinde: Kuhn HW, Tucker AW (editörler). Contributions to the Theory of Games II. Princeton, Princeton University Press, 1953:307-17.
90. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, ve ark. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56-67.
91. Spytek M, Krzyżiński M, Langbein SH, Baniecki H, Wright MN, Biecek P. survex: an R package for explaining machine learning survival models. Bioinformatics. 2023;39(12):btad723.
92. Baniecki H, Kretowicz W, Piatyszek P, Wiśniewski J, Biecek P. dalex: responsible machine learning with interactive explainability and fairness in Python. J Mach Learn Res. 2021;22(214):1-7.
93. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, ve ark. Assessing the performance of prediction models. Epidemiology. 2010;21(1):128-38.
94. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Stat Med. 1999;18(17-18):2529-45.
95. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis. Stat Med. 2004;23(13):2109-23.
96. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med. 2011;30(10):1105-17.
97. Han J, Kamber M, Pei J. Data Mining, 3. Baskı. Waltham, Morgan Kaufmann, 2012:1-740.
98. Gerds TA, Schumacher M. Consistent estimation of the expected Brier score in general survival models. Biom J. 2006;48(6):1029-40.
99. Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. J Am Stat Assoc. 2007;102(478):527-37.
100. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times. Stat Med. 2013;32(30):5381-97.
101. Pachón-García C, Hernández-Pérez C, Delicado P, Vilaplana V. SurvLIMEpy: a Python package implementing SurvLIME. Expert Syst Appl. 2024;237:121620.

102. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th IJCAI*. 1995:1137-43.
103. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD*. 2019:2623-31.
104. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York, Chapman and Hall, 1993:1-436.
105. Spooner A, Chen E, Sowmya A, Sachdev P, Kochan NA, Trollor J, ve ark. A comparison of machine learning methods for survival analysis of high-dimensional clinical data. *Sci Rep*. 2020;10(1):20410.
106. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Sci Rep*. 2021;11(1):6968.
107. Adatorwovor R, Ogunsanya ME, Huang B, Charnigo R, Abraham O. Comparison of Machine Learning Models for Colon Cancer Survival: Predictive Modeling Approach. *JMIR Cancer*. 2025;11:e72665.
108. Moertel CG, Fleming TR, Macdonald JS, ve ark. Levamisole and fluorouracil. *N Engl J Med*. 1990;322(6):352-8.
109. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509.
110. Beran R. Nonparametric regression with randomly censored survival data. Technical Report, University of California, Berkeley; 1981.
111. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*. 2005;92(4):965-70.
112. Steingrimsson JA, Diao L, Strawderman RL. Censoring unbiased regression trees and ensembles. *J Am Stat Assoc*. 2019;114(525):370-83.
113. Ekşi M, Evren İ, Akkaş F, Arıkan Y, Özdemir O, Özlü DN, ve ark. Machine learning algorithms can more efficiently predict biochemical recurrence after robot-assisted radical prostatectomy. *Prostate*. 2021;81(12):913-20.
114. Oosterhoff JHF, de Hond AAH, Peters RM, van Steenberg LN, Sorel JC, Zijlstra WP, ve ark. Machine learning did not outperform conventional competing risk modeling to predict revision arthroplasty. *Clin Orthop Relat Res*. 2024;482(8):1472-82.

115. Kim H, Park T, Jang J, Lee S. Comparison of survival prediction models for pancreatic cancer: Cox model versus machine learning models. *Genomics Inform.* 2022;20(2):e23.
116. Hounye AH, Xiong L, Hou M. Integrated explainable machine learning and multi-omics analysis for survival prediction in cancer with immunotherapy response. *Apoptosis.* 2024;30(1-2):364-88.
117. Stone P, Buckle P, Dolan R, Feliu J, Hui D, Laird BJA, et al. Prognostic evaluation in patients with advanced cancer in the last months of life: ESMO Clinical Practice Guideline. *ESMO Open.* 2023;8(2):101195.
118. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn.* 2006;63(1):3-42.
119. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI Magazine.* 1996;17(3):37-54.
120. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol.* 2015;44(1):324-333.
121. Dunn OJ. Multiple comparisons among means. *J Am Stat Assoc.* 1961;56(293):52-64.
122. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat.* 1979;6(2):65-70.

EKLER

EK-1. Özgeçmiş



EK-2. Etik Kurul Gerekmediğine Dair Yazı

