PAMUKKALE ÜNİVERSİTESİ
Mühendislik Bilimleri Dergisi

PAMUKKALE UNIVERSITY
Journal of Engineering Sciences

Araştırma makalesi | Research paper

Cilt | Vol: XX • Sayı | No: X • ss | pp: XXXX-XXXX • Yıl | Year: 20XX

Authorship attribution in Turkish short fiction: A comparative analysis of linguistic and computational feature sets

Özkan Aslan

ARTICLE INFO

Dates:
Received: 27.04.2026
Accepted: 02.07.2026

Doi:
10.65206/pajes.1938741

Corresponding author:
Özkan Aslan
(oaslan@aku.edu.tr)

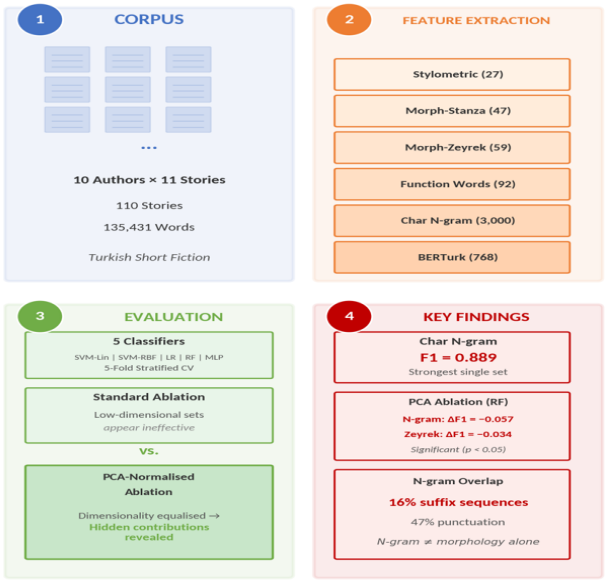
Author addresses:
Department of Computer Engineering, Engineering Faculty, Afyon Kocatepe University, Afyonkarahisar, 03200, Türkiye

ABSTRACT

Context—Authorship attribution is the process of identifying the author of a text based on stylistic features. While methods developed for non-agglutinative languages such as English have matured, agglutinative languages like Turkish remain underexplored. The agglutinative structure of Turkish creates a distinct feature space that is not addressed by English-centric methods.

Objective—This study systematically compares six feature sets across five classifiers for authorship attribution in Turkish short stories. The feature sets are stylometric, morphological (Stanza UD and Zeyrek), function words, character n-grams (TF-IDF), and BERTurk sentence embeddings. To prevent dimensionality imbalance across feature sets from distorting ablation results, a PCA-based dimension-equalized ablation framework is proposed.

Method—A balanced corpus of 110 short stories (135431 words) was compiled from 11 stories by each of 10 authors. Six feature sets were extracted and five classifiers (SVM-Linear, SVM-RBF, Logistic Regression, Random Forest, MLP) were evaluated using 5-fold stratified cross-validation. Standard and PCA-based dimension-equalized ablation analyses were conducted to compare pairwise feature set combinations, and the statistical significance of differences was assessed using paired bootstrap testing. Additionally, n-gram morphological overlap analysis and robustness tests (minimum word count filtering and text truncation) were performed.



Results—Character n-grams achieved the highest individual set performance (F1 = 0.889 ± 0.049, Logistic Regression); BERTurk ranked second with 0.783 ± 0.081. Morphological features reached 0.577 with Stanza and 0.491 with Zeyrek. Standard ablation showed that most feature sets did not provide additional contribution when combined with character n-grams. Under PCA-based dimension-equalized ablation evaluated with repeated cross-validation (10×5), character n-grams made a robust, statistically significant contribution ($\Delta F1 \approx 0.07$, $p < 0.001$), and low-dimensional stylometric and function-word feature sets also contributed significantly once dimensionality was equalized; in contrast, the morphological feature sets showed no contribution that survived the higher-powered test. N-gram morphological overlap analysis showed that 16% of the most discriminative n-grams correspond to Turkish suffix sequences.

Conclusion—The apparent ineffectiveness of morphological and stylometric features in standard ablation partly reflects dimensionality imbalance across feature sets. Under PCA-based dimension-equalized ablation with repeated cross-validation, character n-grams and the low-dimensional stylometric and function-word sets contribute significantly, whereas the morphological sets do not survive the higher-powered test; their apparent effect in a single 5-fold run did not reproduce. Ablation designs ignoring dimensionality imbalance in agglutinative languages may thus mislead, and significance claims on small corpora warrant caution. Future work includes expanding the corpus, addressing author-topic confounding, and testing other agglutinative languages.

Key Words—Authorship attribution, Turkish NLP, Character n-grams, Morphological features, Dimension-equalized ablation, BERTurk.

PAMUKKALE ÜNİVERSİTESİ
Mühendislik Bilimleri Dergisi

PAMUKKALE UNIVERSITY
Journal of Engineering Sciences

Araştırma makalesi | Research paper

Cilt | Vol: XX • Sayı | No: X • ss | pp: XXXX-XXXX • Yıl | Year: 20XX

Türkçe kısa öykülerde yazar tanıma: dilbilimsel ve hesaplamalı özellik setlerinin karşılaştırmalı analizi

Özkan Aslan

MAKALE BİLGİLERİ

Tarihler:
Geliş: 27.04.2026
Kabul: 02.07.2026

Doi:
10.65206/pajes.1938741

Sorumlu yazar:
Özkan Aslan
(oaslan@aku.edu.tr)

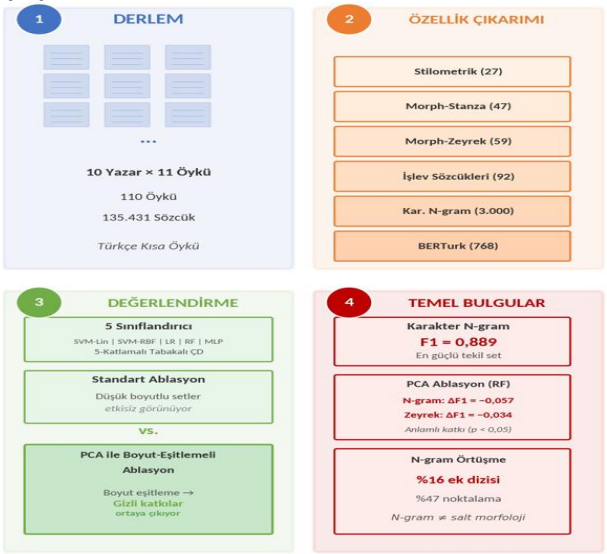
Yazar adresleri:
Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Afyon Kocatepe Üniversitesi, Afyonkarahisar, 03200, Türkiye

ÖZ

Arka Plan—Yazar tanıma (authorship attribution) bir metnin üslup özelliklerine dayanarak yazarını belirleme sürecidir. İngilizce gibi eklemeli (bitişken) olmayan diller için geliştirilen yöntemler olgunlaşmış olsa da Türkçe gibi sondan eklemeli diller yeterince araştırılmamıştır. Türkçenin bu yapısı İngilizce merkezli yöntemlerin kapsamadığı farklı bir özellik uzayı oluşturmaktadır.

Amaç—Bu çalışma Türkçe kısa öykülerde yazar tanıma amacıyla altı farklı özellik setini beş sınıflandırıcı model üzerinde sistematik olarak karşılaştırmaktadır. Bu özellik setleri stilometrik, morfolojik (Stanza UD ve Zeyrek), işlev sözcükleri, karakter n-gramları (TF-IDF) ve BERTurk cümle vektörleridir (sentence embeddings). Özellik setleri arasındaki boyut dengesizliğinin ablasyon sonuçlarını çarpıtmasını önlemek için PCA ile boyut-eşitlemeli bir ablasyon çerçevesi önerilmektedir.

Yöntem—On yazardan 11'er kısa öykü ile dengeli bir derlem oluşturulmuştur (110 öykü, 135431 sözcük). Altı özellik seti çıkarılmış ve beş sınıflandırıcı (SVM-Doğrusal, SVM-RBF, Lojistik Regresyon, Rastgele Orman, Çok Katmanlı Algılayıcı) 5 katlı tabakalı çapraz doğrulama ile değerlendirilmiştir. Standart ve PCA ile boyut-eşitlemeli ablasyon analizleri ile ikili özellik seti kombinasyonları karşılaştırılmış, farkların istatistiksel anlamlılığı eşlenmiş bootstrap testi ile sınanmıştır. Ayrıca n-gram morfolojik örtüşme analizi ve sağlamlık sınamaları (minimum sözcük sayısı filtreleme ve metin sınırlandırma) gerçekleştirilmiştir.



Bulgular—Karakter n-gramları en yüksek tekil set performansına ulaşmıştır (F1 = 0.889 ± 0.049, Lojistik Regresyon); BERTurk 0.783 ± 0.081 ile ikinci sıradadır. Morfolojik özellikler Stanza ile 0.577, Zeyrek ile 0.491 değerine ulaşmıştır. Standart ablasyon, çoğu özellik setinin karakter n-gramları ile birleştirildiğinde ek katkı sağlamadığını göstermiştir. Ancak PCA ile boyut-eşitlemeli ablasyon, tekrarlı çapraz doğrulama (10×5) ile değerlendirildiğinde, karakter n-gramlarının sağlam ve istatistiksel olarak anlamlı bir katkı sağladığını ($\Delta F1 \approx 0.07$, $p < 0.001$); boyut eşitlendiğinde düşük-boyutlu stilometrik ve işlev sözcüğü setlerinin de anlamlı katkı verdiğini göstermiştir. Buna karşılık morfolojik setler yüksek-güçlü testte korunan bir katkı göstermemiştir. Bu sonuçlar özellik setleri arasında belirgin boyut farkı olduğunda standart ablasyonun yanıltıcı sonuçlar üretebileceğini göstermektedir. N-gram morfolojik örtüşme analizi en ayırt edici n-gramların %16'sının Türkçe ek dizilerine karşılık geldiğini ortaya koymuştur.

Sonuç—Standart ablasyonda morfolojik ve stilometrik özelliklerin etkisiz görünmesi kısmen özellik setleri arasındaki boyut dengesizliğini yansıtır. PCA ile boyut-eşitlemeli ablasyon, tekrarlı çapraz doğrulama ile değerlendirildiğinde, karakter n-gramları ile düşük-boyutlu stilometrik ve işlev sözcüğü setleri anlamlı katkı sağlarken morfolojik setler yüksek-güçlü testte korunmamıştır; tek bir 5-katlı koşudaki görünür katkıları yeniden üretilememiştir. Bu nedenle bitişken dillerde boyut dengesizliğini göz ardı eden ablasyon tasarımları yanıltıcı olabilir ve küçük derlemlerde anlamlılık iddiaları temkinli kurulmalıdır. Gelecek çalışmalarda derlemin genişletilmesi, yazar-konu karışmasının ele alınması ve çerçevenin farklı bitişken dillerde sınanması planlanmaktadır.

Anahtar Kelimeler—Yazar tanıma, Türkçe doğal dil işleme, Karakter n-gramları, Morfolojik özellikler, Boyut-eşitlemeli ablasyon, BERTurk.

I. GİRİŞ (INTRODUCTION)

Yazar tanıma (authorship attribution, AA) bir metnin üslup örüntülerine dayanarak yazarını belirlemeye yönelik hesaplamalı bir süreçtir. Adli dilbilim, intihal tespiti ve siber güvenlik gibi çok sayıda uygulama alanına sahip olan bu disiplin [1] ve [2] gibi basit istatistiksel sayımlardan derin öğrenme mimarilerine uzanan geniş bir yöntemsel evrimden geçmiştir [3], [4].

Modern stilometride öne çıkan yaklaşımlar arasında işlev sözcükleri profilleme [5], [6], karakter n-gramları [7], [8], sözcüksel zenginlik ölçümleri (Yule's K, hapax oranı) [1] ve son yıllarda BERT ailesi dönüştürücü tabanlı gömmeler [9] sayılabilir. Bu yöntemler ağırlıklı olarak İngilizce olmak üzere Avrupa dilleri üzerinde geliştirilmiş ve doğrulanmıştır [1], [10]. Rybicki ve Eder (2011) standart Delta yönteminin morfolojik olarak zengin dillerde (Lehçe, Latince) performansının düştüğünü göstermiştir [11]. Delta'nın ayırt edici gücü, düşük ve orta frekanslı sözcüklerin kolektif katkısına dayandığından [12], yüzey sözcük biçimlerinin çekim çeşitliliği nedeniyle seyrelmesi bu performans düşüşünün olası nedenleri arasındadır.

Türkçe sondan eklemeli (bitişken) bir dil olup; tek bir kök üzerine eklenen yapım ve çekim ekleri ile çok yüksek sayıda farklı yüzey biçimi üretilebilmektedir [13]. Bu yapısal özellik doğal dil işleme (DDİ) sistemleri için ciddi zorluklar oluşturmaktadır: bilinmeyen sözcük sorunu (out-of-vocabulary, OOV), veri seyrekliği ve yüksek boyutluluk bunların başında gelmektedir [13]-[16]. Öte yandan Türkçenin morfolojik zenginliği yazar tanıma açısından benzersiz bir özellik uzayı sunmaktadır. Zaman/görünüş ekleri, kanıtsallık işaretleyicileri (-mIş/-DI karşıtlığı) [17], [18], zarf-fiiller (-Ip, -ArAk, -IncA) [19] ve durum ekleri bir yazarın bilinçdışı dilsel tercihlerini yansıtma potansiyeli taşımaktadır [1], [6]. Ayrıca morfo-sentaktik özelliklerin sözcük dağarcığından bağımsız olarak yazar sinyali taşıdığı yüksek doğrulukla gösterilmiştir [20]. Türkçe üzerine gerçekleştirilen yazar tanıma çalışmaları sınırlı sayıdadır. Amasyalı ve Diri (2006) karakter n-gramları ile yazar, tür ve cinsiyet sınıflandırması gerçekleştirmiş [21], Türkoğlu vd. (2007) ise n-gram ve yazar niteliklerinin birleştirilmesiyle en iyi sonuçların elde edildiğini göstermiştir [22]. Bay ve Çelebi (2015) gazete köşe yazılarında %99,7 doğruluk elde etmiştir [23]. Güllü ve Polat (2022) topluluk öğrenmesi ve genetik algoritma tabanlı özellik seçimi ile geniş ölçekli Türkçe yazar tanıma gerçekleştirmiştir [24]. Pirli (2025) 22 yazardan 94 Türkçe roman ve kısa öykü derlemi üzerinde lemma, token, POS etiketi ve karakter n-gramlarını karşılaştırarak lemmaların tokenlarla karşılaştırılabilir sonuçlar verdiğini ve Türkçenin bitişken yapısının sözcüksel tercihleri çekim biçimleri altında gizleyebildiğini bildirmiştir [25]. Saygılı vd. (2016) Türkçe'ye özgü özelliklerden yararlanarak fiilimsi frekanslarının yazar tanımada ayırt edici sinyal taşıdığını belirtmiştir [26]. Gündüz (2025), daha geniş Türkçe metin sınıflandırma ve profilleme bağlamında derin gömmeler ve meta-sezgisel özellik seçimi ile cinsiyet profilleme gerçekleştirmiştir [27]. Yıldırım ve Yıldız (2018) ise geleneksel sözcük torbası tabanlı yaklaşımlar ile sinir ağı temelli sözcük vektörleri arasında sistematik bir karşılaştırma gerçekleştirmiş ve etkili özellik seçimi uygulandığında geleneksel yöntemlerin yeni nesil sözcük vektörleriyle yarışabilir düzeyde kaldığını saptamıştır [28]. Kestemont (2014), morfolojik olarak zengin dillerde dilbilgisel işlevlerin serbest biçimbirimler yerine bağlı biçimbirimler (suffix) aracılığıyla ifade edildiğini vurgulamıştır [6]. Bu gözlem karakter n-gramlarının bitişken dillerde neden güçlü olduğuna dair kuramsal bir çerçeve sunar: n-gramlar, ek dizilerini dolaylı olarak yakalayarak morfolojik bilgiye erişebilmektedir [7]. Sapkota vd. (2015) ayrıca karakter n-gramlarının ayırt edici gücünün homojen olmadığını, ek (affix) ve noktalama ile ilişkili n-gramların en güçlü alt kategoriler olduğunu tespit etmiştir [7].

Mevcut çalışmaların önemli bir metodolojik sınırlılığı, farklı boyutlardaki özellik setlerinin ablasyon yoluyla karşılaştırılmasında boyut dengesizliğinin dikkate alınmamasıdır. Yüksek boyutlu özellik setleri düşük boyutlu setlerle birleştirildiğinde düşük boyutlu setlerin etkisi bastırılabilmekte ve bu durum yanıltıcı sonuçlara yol açabilmektedir. Ayrıca karakter n-gramlarının morfolojik olarak zengin dillerde neden güçlü performans gösterdiği (morfolojik bilgiyi dolaylı olarak yakalayıp yakalamadığı) deneysel olarak yeterince araştırılmamıştır. Bu çalışmanın katkıları şu şekilde özetlenebilir: (*i*) Türkçe kısa öyküler üzerinde altı farklı özellik setini (iki farklı morfolojik çözümleyici dahil) sistematik olarak karşılaştırmaktadır; (*ii*) PCA ile boyut-eşitlemeli ablasyon çerçevesi önerilerek boyut dengesizliğinin ablasyon sonuçlarını nasıl etkilediği deneysel olarak gösterilmiştir; (*iii*) en ayırt edici karakter n-gramlarının morfolojik ek dizileriyle örtüşme oranı ölçülmüştür; (*iv*) morfolojik özelliklerin sınıflandırma performansı düşük olsa da yazar sinyali için dilbilimsel yorumlanabilirlik sunduğu saptanmıştır.

II. YÖNTEM (METHOD)

Bu bölümde; veri seti, özellik çıkarım yöntemleri, sınıflandırıcı modeller ve değerlendirme çerçevesi açıklanmaktadır.

A. Veri seti (Dataset)

Çalışmada on Türk yazardan oluşan dengeli bir kısa öykü derlemi kullanılmıştır. Her yazar 11 kısa öykü ile temsil edilmekte olup derlem toplamda 110 öykü ve 135431 sözcük içermektedir. Yazarlar Türk edebiyat tarihinin farklı dönemlerini ve üslup eğilimlerini temsil edecek şekilde seçilmiştir [29], [30]: Sait Faik Abasıyanık, Hüseyin Rahmi Gürpınar, Tomris Uyar, Aziz Nesin, Ömer Seyfettin, Nazım Hikmet, Nezihe Meriç, Sevgi Soysal, Orhan Kemal ve Yusuf Atılgan. Her öykü cümlelere ayrılmış ve JSON formatında depolanmıştır.

Öyküler arası sözcük sayısı önemli farklılıklar göstermektedir (minimum 197, maksimum 11266 sözcük). Eder (2015), güvenilir yazar tanıma için minimum 5.000 sözcüklük örnekler önermiş olmakla birlikte bu eşik roman ve uzun metin çalışmaları için geçerlidir; kısa öykü türü doğası gereği daha kısa metinlerden oluşmaktadır [31]. Bu dengesizliğin ve kısa metin uzunluğunun sonuçlar üzerindeki etkisi Bölüm III.F'de sağlamlık sınaması aracılığıyla değerlendirilmektedir.

Yazarlar; (*i*) modern Türk öykücülüğünün başlıca dönem ve üslup eğilimlerini kapsayacak, (*ii*) tür kısa öyküyle sınırlı tutulacak, (*iii*) her yazardan en az 11 öykü ile dengeli bir tasarım sağlanacak ve (*iv*) güvenilir, yayımlanmış derlemelerden erişilebilecek biçimde seçilmiştir. Yazarların bir kısmı telif kapsamında olduğundan öykülerin ham metinleri paylaşılmamaktadır; öykülerin alındığı yayımlanmış baskılar Ek-C'de listelenmiştir.

B. Özellik çıkarımı (Feature extraction)

1. Stilometrik özellikler (Stylometric features)

Cümle uzunluğu ortalaması ve standart sapması (sözcük sayısı), sözcük uzunluğu ortalaması ve standart sapması (karakter sayısı), type-token oranı (TTR), hapax legomena oranı, Yule's K sabiti ve 20 farklı noktalama işareti frekansı (sözcük başına normalize edilmiş) olmak üzere toplam 27 stilometrik özellik çıkarılmıştır. Tüm özellikler düzenli ifade (regex) tabanlı Türkçe tokenizör ile tokenize edilmiş metinler üzerinden hesaplanmıştır.

TTR, bir metindeki sözcük çeşitliliğini ölçer [32] ve farklı sözcük sayısının (V) toplam sözcük sayısına (N) bölünmesiyle elde edilir ($TTR = V/N$). Hapax legomena oranı metin içinde yalnızca bir kez geçen sözcüklerin toplam sözcük sayısına oranıdır ve yazarın sözcük dağarcığının genişliğine işaret eder.

Yule's K sabiti sözcüksel zenginliğin metin uzunluğundan bağımsız bir ölçüsüdür [33]. i frekansında kullanılan sözcük sayısı $f(i, N)$ ve toplam sözcük sayısı N olmak üzere (1)'deki gibi hesaplanır:

$$K = 10^4(\sum i^2 f(i,N) - N)/N^2. \quad (1)$$

Yüksek K değeri sözcüklerin dar bir kümede yoğunlaştığını (düşük sözcüksel çeşitlilik), düşük K değeri ise sözcüklerin daha geniş bir dağarcığa yayıldığını gösterir.

2. Morfolojik özellikler: Stanza UD (Morphological features: Stanza UD)

Morfolojik özellik çıkarımı için Stanza kütüphanesinin [34] Türkçe Evrensel Bağımlılıklar (UD) hattı kullanılmıştır. Bu hat; tokenizasyon, çok-sözcüklü token açımı, POS etiketleme, morfolojik özellik etiketleme, lemmatizasyon ve bağımlılık ayrıştırma işlemlerini içermektedir. Her öykünün tam metni Stanza'dan geçirilmiş ve 47 morfolojik özellik çıkarılmıştır: 14 UPOS etiketi oranı (toplam token sayısına normalize edilmiş), 14 bağımlılık ilişkisi oranı (nsubj, obj, obl, advmod, amod, nmod, conj, compound, flat, parataxis, advcl, acl, ccomp, xcomp), fiil sayısına normalize edilmiş 5 zaman/görünüş oranı (present, future, perfective, imperfective, progressive), 4 kanıtsallık özelliği (Evident = Nfh etiketi ile belirlenen -mIş ve yalnızca Tense = Past ile belirlenen -DI [17] için hem toplam geçmiş zaman içindeki oran hem de fiil sayısına normalize edilmiş sıklık), 7 durum eki dağılımı (Nom, Acc, Dat, Loc, Abl, Gen, Ins), VerbForm=Conv etiketli tokenların fiil sayısına oranı olarak zarf-fiil oranı [19], fiil/isim oranı ve yan cümle indeksi (subordination index: advcl, acl, ccomp ve xcomp bağımlılıklarının toplam cümle sayısına oranı).

3. Morfolojik özellikler: Zeyrek (Morphological features: Zeyrek)

İkinci bir morfolojik çözümleyici olarak Zeyrek kütüphanesi [35] kullanılmıştır. Zeyrek, Zemberek morfolojik çözümleyicisinin Python uyarlamasıdır [36] ve UD etiketleri yerine doğrudan ek dizilerini çıkarmaktadır. Öncelikle derlem taranarak en sık 40 ek etiketi (suffix tag) ve 15 ek ikilisi (suffix bigram) keşfedilmiştir. Bu keşif tüm derlem üzerinden gerçekleştirilmiş olup; örnekleme dayalı bir özellik seçimi değil, dilde mevcut ek etiketlerinin envanterini belirleyen bir ön adımdır. Her öykü için 40 ek etiketi frekansı (sözcük sayısına normalize edilmiş), ortalama ek sayısı, ekli sözcük yüzdesi, ortalama ve standart sapma kök/sözcük oranı ve 15 ek ikilisi frekansı çıkarılmıştır. Toplam özellik sayısı 59'dur. İki çözümleyicinin karşılaştırılması, morfolojik özellik çıkarımının araca bağımlılığını değerlendirmeyi amaçlamaktadır.

4. İşlev sözcükleri (Function words)

İşlev sözcüklerinin yazar tanımadaki kuramsal dayanağına [6] uygun olarak Türkçeye özgü 92 işlev sözcüğü (bağlaçlar, edatlar, zamirler, belirteçler, soru ekleri, söylem belirteçleri) listesi oluşturulmuş ve her öyküdeki frekansları toplam sözcük sayısına normalize edilerek hesaplanmıştır.

5. Karakter n-gramları (Character n-grams)

İki-dört karakter aralığında TF-IDF karakter n-gramları çıkarılmıştır (sklearn TfidfVectorizer, analyzer='char', sublinear_tf=True). Kelime sınırı kısıtlaması uygulanmadığından kelimeler arası geçişleri içeren n-gramlar da dahil edilmiştir. Veri sızıntısını önlemek amacıyla TfidfVectorizer her çapraz doğrulama katlamasında yalnızca eğitim verileri üzerinde kurulmuş ve test verilerine yalnızca dönüşüm uygulanmıştır. Maksimum özellik sayısı 3000 olarak sınırlandırılmıştır. Karakter n-gram aralığının (n = 2–4) seçimini gerekçelendirmek için farklı aralıklar aynı katlama-içi TF-IDF ve çapraz doğrulama protokolüyle karşılaştırılmıştır (Tablo 1). En iyi başarı, trigram içeren aralıklarda elde edilmiştir; (2, 4) aralığı en iyi sonuca çok yakın (≈ 0.03 F1 fark, örtüşen güven aralıkları) olup hem bigram düzeyindeki noktalama/sınır ipuçlarını hem de trigram düzeyindeki morfolojik ipuçlarını birlikte yakaladığı için tercih edilmiştir.

Tablo 1. Karakter n-gram aralığı taraması (en iyi model F1).

Table 1. *Character n-gram range sweep (best model F1).*

Aralık (n)	En iyi model F1
2–2	0.833
3–3	0.916
4–4	0.899
2–3	0.897
2–4	0.889
2–5	0.862
3–5	0.868

6. BERTurk cümle vektörleri (BERTurk sentence embeddings)

Her öykünün cümleleri dbmdz/bert-base-turkish-cased modeli [37] kullanılarak kodlanmıştır. Her cümle için token düzeyinde ortalama havuzlama uygulanmış, ardından tüm cümle vektörleri ortalaması alınarak 768 boyutlu bir öykü vektörü elde edilmiştir. Maksimum token uzunluğu 256 alt sözcük olarak belirlenmiştir.

BERTurk temsilinin havuzlama seçimine duyarlılığını incelemek için ortalama (mean), [CLS] simgesi ve maksimum (max) havuzlama aynı koşullarda karşılaştırılmış; en iyi başarılar sırasıyla max 0.805, ortalama 0.783 ve [CLS] 0.691 olmuştur. Ayrıca güncel bir Türkçe cümle-dönüştürücü modeli de [38] değerlendirilmiş, F1 = 0.582 ile en düşük değerde kalmıştır. Tüm transformer tabanlı temsiller karakter n-gramlarının (0,889) altında kalmıştır; cümle-dönüştürücünün görece düşük başarısı, bu modellerin anlamsal benzerlik (içerik) için eniyilenmiş olmasıyla tutarlıdır; oysa yazar tanımada ayırt edici olan içerik değil üsluptur.

C. Sınıflandırıcı modeller (Classifier models)

Beş sınıflandırıcı model karşılaştırılmıştır: (*i*) SVM-Doğrusal (C=1.0), (*ii*) SVM-RBF (C=10.0, gamma=scale), (*iii*) Lojistik Regresyon (LogReg; solver=lbfgs, multinomial, C=1.0), (*iv*) Rastgele Orman (RF; 300 ağaç), (*v*) Çok Katmanlı Algılayıcı (ÇKA/MLP; 256-128-64 nöron, erken durdurma %15 validasyon oranıyla, maks. 500 iterasyon). Tüm modellerde class_weight='balanced' parametresi kullanılmıştır. Her katlamada StandardScaler ile özellikler ölçeklendirilmiştir.

D. Değerlendirme çerçevesi (Evaluation framework)

Değerlendirme, 5 katlı tabakalı çapraz doğrulama (5-fold stratified CV) ile gerçekleştirilmiştir. Birincil metrik F1 makro olup doğruluk ikincil metrik olarak raporlanmıştır.

Dört farklı analiz yürütülmüştür: (*i*) her seferinde bir özellik setinin çıkarıldığı standart ablasyon çalışması, (*ii*) tüm özellik setlerinin PCA ile eşit boyuta (50 bileşen veya orijinal boyut, hangisi küçükse) indirildikten sonra yapılan PCA ile boyut-eşitlemeli ablasyon, (*iii*) tüm ikili kombinasyonların test edildiği analiz ve (*iv*) sağlamlık sınaması (≥ 500 sözcük alt kümesi ve 2000 sözcüğe sınırlandırma). PCA her çapraz doğrulama katlamasında yalnızca eğitim verisi üzerinde hesaplanmış olup test verisine yalnızca dönüşüm uygulanmıştır.

İstatistiksel anlamlılık fold düzeyinde F1 skorları üzerinde eşlenmiş bootstrap testi (10000 örneklem) ile değerlendirilmiştir.

N-gram morfolojik örtüşme analizi için, tam derlem üzerinde eğitilmiş Lojistik Regresyon modelinin en yüksek mutlak katsayıya sahip 200 karakter n-gramı seçilmiş ve her biri Türkçe morfolojik ek listesiyle eşleştirilerek ek ilişkili (suffix), sözcük sınırı (word-boundary), noktalama (punctuation) ve kök/diğer (root/other) olarak kategorize edilmiştir.

Tekrarlanabilirlik için tüm rastgele işlemlerde (tabakalı çapraz doğrulama katlama ayrımı, sınıflandırıcı başlatma, PCA, bootstrap yeniden örnekleme) sabit tohum değeri (42) kullanılmıştır. Kullanılan yazılım sürümleri şöyledir: Stanza v1.11.1, Zeyrek v0.1.3 ve BERTurk (dbmdz/bert-base-turkish-cased) [37]; Türkçe cümle-dönüştürücü olarak [38] modeli kullanılmıştır.

Boyut-eşitlemeli ablasyonda kullanılan PCA bileşen sayısının seçimi, 20–100 bileşen aralığında duyarlılık analiziyle

gerekçelendirilmiştir; en iyi F1 değerleri 20→0.845, 30→0.848, 50→0.873, 75→0.855 ve 100→0.861 olup en yüksek başarı 50 bileşende elde edilmiştir. 50 bileşen düşük-boyutlu setlerde varyansın tamamına yakınını, İşlev Sözcükleri ve BERTurk setlerinde ise sırasıyla %92.7 ve %92.6'sını korumaktadır. Bu nedenle tüm boyut-eşitlemeli analizlerde 50 bileşen kullanılmıştır.

III. BULGULAR ve TARTIŞMA (RESULTS and DISCUSSION)

Bu bölümde tekil özellik seti karşılaştırması, ablasyon çalışması, n-gram morfolojik örtüşme analizi, morfolojik özelliklerin yorumlanabilirlik değeri, ikili kombinasyon analizi, sağlamlık sınaması, yazar bazlı performans ve sınırlılıklar sunulmaktadır.

A. Tekil özellik seti karşılaştırması (Single feature set comparison)

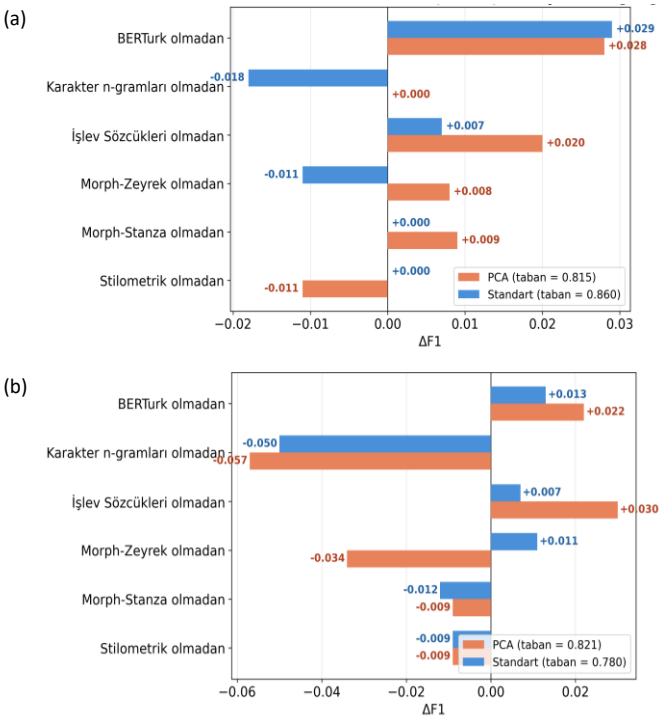
Tablo 2 her özellik seti için en iyi performansı gösteren model ve F1 makro sonuçlarını sunmaktadır.

Karakter n-gramları (TF-IDF) en yüksek tekil set performansına ulaşmıştır (F1 = 0.889). Bu sonuç karakter n-gramlarının stilometride güçlü bir temel oluşturduğuna dair uluslararası literatürle [7], [8] tutarlıdır. Karakter n-gramlarının BERTurk'e üstünlüğü istatistiksel olarak anlamlıdır (Δ = 0.106, bootstrap p = 0.023); ayrıca birleşik modelden de anlamlı biçimde yüksektir (Δ = 0.029, p = 0.005). BERTurk vektörleri 0.783 ile ikinci sırada yer almıştır. Stilometrik özellikler 0.682 ile üçüncü sıradadır. Morfolojik özellik setleri daha düşük performans göstermiştir: Stanza 0.577 ve Zeyrek 0.491. Stanza ile Zeyrek arasındaki fark istatistiksel olarak anlamlı bulunmamıştır (bootstrap p = 0.084); benzer şekilde Stanza ile stilometrik set arasındaki fark da anlamlı değildir (p = 0.155). Buna karşılık, Stanza'nın BERTurk'ten düşük performansı anlamlıdır (Δ = −0.206, p < 0.001). Bu iki morfolojik çözümleyicinin benzer düzeyde performans üretmesi, morfolojik özelliklerin sınıflandırma gücünün araca özgü değil, genel olarak sınırlı olduğuna işaret etmektedir. Tüm özelliklerin birleştirilmesi (F1 = 0.860) karakter n-gramları tekil setini (0.889) aşamamıştır. Bu durumun olası nedenleri Bölüm III.B'de ablasyon çerçevesinde tartışılmaktadır.

Sınıflandırıcı modeller açısından LogReg, karakter n-gramları ve birleşik model için en iyi performansı gösterirken RF, düşük boyutlu setlerde (stilometrik, morfolojik) daha tutarlı sonuçlar üretmiştir. Tüm özellik seti × sınıflandırıcı kombinasyonlarının ayrıntılı sonuçları Ek-A'da sunulmaktadır.

B. Ablasyon çalışması ve PCA ile boyut-eşitlemeli ablasyon (Ablation study and PCA-based dimension-equalized ablation)

Şekil 1 standart ve PCA ile boyut-eşitlemeli ablasyon sonuçlarını aynı sınıflandırıcı üzerinden karşılaştırmaktadır. Standart ablasyon ile PCA ablasyonunun farklı sınıflandırıcılarla değerlendirilmesi gözlenen farkın boyut eşitlemeye mi yoksa sınıflandırıcı değişimine mi ait olduğu sorusunu gündeme getirebilir. Bu karışıklığı önlemek için her iki ablasyon türü hem LogReg hem RF ile ayrı ayrı uygulanmıştır. Veri sızıntısını



Şekil 1. Standart ve PCA ile boyut-eşitlemeli ablasyon karşılaştırması: (a) Lojistik Regresyon, (b) Rastgele Orman.

Figure 1. *Standard vs. PCA-based dimension-equalized ablation comparison: (a) Logistic Regression, (b) Random Forest.*

önlemek için, Bölüm II.D'de belirtildiği üzere, PCA her katlamada yalnızca eğitim verisi üzerinde hesaplanmıştır.

Standart ablasyonda (LogReg, taban F1 = 0.860): stilometrik, Morph-Stanza, işlev sözcüğü ve BERTurk setlerinin çıkarılması F1 değerinde sıfır veya pozitif değişime neden olmuştur. Yalnızca Morph-Zeyrek (ΔF1 = −0.011) ve karakter n-gramları (ΔF1 = −0.018) çıkarıldığında küçük düşüşler gözlenmiştir. Bu sonuçlar birleşik modeldeki sinyalin büyük ölçüde karakter n-gramları tarafından taşındığını düşündürmektedir. Öte yandan BERTurk'ün birleşik modelden çıkarılması performansı artırmıştır (ΔF1 = 0.029); bu durum yüksek boyutlu özellik setlerinin birleştirilmesinde gürültü birikiminin performansı olumsuz etkileyebileceğine işaret etmektedir. Sinyalin büyük ölçüde karakter n-gramlarında yoğunlaştığı bulgusu RF sınıflandırıcısında da desteklenmiştir: yalnızca karakter n-gramlarının çıkarılması belirgin bir düşüşe (ΔF1 = −0.050) neden olmuştur.

PCA ile boyut-eşitlemeli ablasyon sonuçları sınıflandırıcıya göre farklılaşmıştır. LogReg için PCA ablasyonu standart ablasyonla benzer bir örüntü sergilemiştir: boyut eşitleme, düşük boyutlu setlerin etkisini belirgin biçimde görünür kılmamıştır. RF için ise karakter n-gramlarının çıkarılması en büyük kaybı üretmiştir (ΔF1 = −0.057); Morph-Zeyrek'in çıkarılması da kayda değer bir düşüşe neden olmuştur (ΔF1 = −0.034, bootstrap p = 0.020). Bu bulgular PCA ablasyonunun bazı özellik setlerinin olumlu etkisini

Tablo 2. Tekil özellik seti karşılaştırması (en iyi model).

Table 2. *Single feature set comparison (best model).*

Özellik Seti	Boyut	Model*	F1**	Doğruluk
TF-IDF***	maks. 3000	LogReg	0.889 ± 0.049	0.900 ± 0.034
BERTurk	768	SVM-Linear	0.783 ± 0.081	0.791 ± 0.074
Stilometrik	27	RF	0.682 ± 0.050	0.700 ± 0.055
Morph-Stanza	47	RF	0.577 ± 0.123	0.591 ± 0.111
İşlev Sözcükleri	92	RF	0.546 ± 0.070	0.582 ± 0.073
Morph-Zeyrek	59	RF	0.491 ± 0.136	0.518 ± 0.131
Tümü	3993	LogReg	0.860 ± 0.052	0.873 ± 0.045

* En iyi model, ** F1 makro, *** Karakter n-gramları

görünür kılabildiğini ancak etkinin sınıflandırıcı seçimine bağlı olduğunu göstermektedir.

Ablasyon deltalarının istatistiksel anlamlılığı, hem tek 5-katlı hem de tekrarlı (10×5 = 50 skor) çapraz doğrulama ile eşli bootstrap kullanılarak değerlendirilmiştir (Tablo 3). Tekrarlı çapraz doğrulamada, RF altında karakter n-gramlarının çıkarılması en büyük ve sağlam kaybı üretmiştir (ΔF1 = 0.070; %95 GA [0.047, 0.093]; $p < 0.001$); düşük-boyutlu stilometrik (ΔF1 = 0.030; $p = 0.019$) ve işlev sözcüğü setlerinin katkıları da anlamlı bulunmuştur. Buna karşılık morfolojik setlerin (Morph-Zeyrek, Morph-Stanza) çıkarılması anlamlı bir kayba yol açmamıştır (RF için $p \approx 0.6$–0.8). Tek bir 5-katlı koşuda Morph-Zeyrek için gözlenen sınırda anlamlılık ($p \approx 0.02$) tekrarlı çapraz doğrulamada korunmamış olup, küçük örneklemli bootstrap'in düşük gücüne atfedilmektedir. Tüm deltalar ΔF1 ve %95 güven aralıklarıyla raporlanmıştır. Sonuçlar standart ablasyonda morfolojik özelliklerin birleşik modelde etkisiz görünmesinin kısmen boyut dengesizliğiyle açıklanabileceğine ancak PCA eşitlemenin etkisinin sınıflandırıcıya bağlı olduğuna işaret etmektedir. RF sınıflandırıcısı PCA ablasyonunda daha belirgin düşüşler üretirken (özellikle karakter n-gramları ve Morph-Zeyrek için), LogReg sınıflandırıcısı her iki koşulda da benzer örüntüler sergilemiştir. Bu bulgu boyut eşitlemenin tek başına yeterli olmadığını ve sınıflandırıcı seçiminin ablasyon yorumlarını doğrudan etkilediğini ortaya koymaktadır.

Bu bulguların uluslararası literatürdeki ablasyon sonuçlarıyla karşılaştırılması önemlidir. İngilizce üzerinde Sapkota vd. [7], affix ve noktalama tabanlı alt n-gramların bütün karakter n-gramı kümesinin performansını tek başına yakaladığını, sözcük tabanlı n-gramların çıkarılmasının doğruluğu azaltmadığını göstermiştir. Türkçe derlemimizde ise PCA ile boyut-eşitlemeli ablasyon (RF) çerçevesinde karakter n-gramlarının yanı sıra Morph-Zeyrek özelliklerinin çıkarılması da istatistiksel olarak anlamlı kayıplara yol açmıştır. Bu fark bitişken bir dilde açık morfolojik çözümlemenin karakter düzeyi özelliklerden bağımsız ek bir sinyal taşıyabileceğini düşündürmektedir.

C. N-gram morfolojik örtüşme analizi (N-gram morphological overlap analysis)

Karakter n-gramlarının neden güçlü olduğunu anlamak için Lojistik Regresyon modelinin en ayırt edici 200 n-gramı kategorize edilmiştir. Buna göre noktalamalarla ilişkili n-gramlar baskın çıkmıştır (94/200, %47). Bunu kök/diğer (44/200, %22), sözcük sınırı (30/200, %15), ek ilişkili (23/200, %11.5) ve ek + sınır (9/200, %4.5) kategorileri izlemiştir. Ek ilişkili toplam (suffix + suffix+boundary) 32/200 (%16) olup beklentinin altında kalmıştır. Bu sonuç karakter n-gramlarının gücünün yalnızca morfolojik ek dizilerini yakalamasıyla açıklanamayacağını göstermektedir. Noktalama tercihleri (virgül, nokta, ünlem işareti kullanım sıklığı ve bağlamı) ve sözcük sınırı örüntüleri de güçlü yazar sinyalleri taşımaktadır. Bu bulgu Sapkota vd.'nin [7] affix ve noktalama n-gramlarının en güçlü alt kategoriler olduğu sonucuyla tutarlıdır. Bununla birlikte, otomatik kategorilemenin sınırlılıkları göz önünde bulundurulmalıdır. Kısa n-gramlar (2 karakter) birden fazla kategoriye ait olabilir; örneğin "de" hem durum eki hem bağlaç hem de bir sözcüğün parçası olabilir. Bu belirsizlik ek ilişkili oranın olduğundan düşük hesaplanmasına neden olmuş olabilir.

Tablo 3. Tekrarlı CV (10×5) ablasyon (RF, PCA).

Table 3. *Repeated CV (10×5) ablation (RF, PCA).*

Çıkarılan set	ΔF1	p
Char n-grams	+0.070	< 0.001
Stylometric	+0.030	0.019
Function Words	+0.017	0.069
BERTurk	+0.019	0.057
Morph-Zeyrek	−0.005	0.63
Morph-Stanza	−0.002	0.81

D. Morfolojik özelliklerin yorumlanabilirlik değeri (Interpretability value of morphological features)

Morfolojik özellikler tekil sınıflandırma performansı açısından sınırlı kalsa da önemli bir yorumlanabilirlik avantajı sunmaktadır.

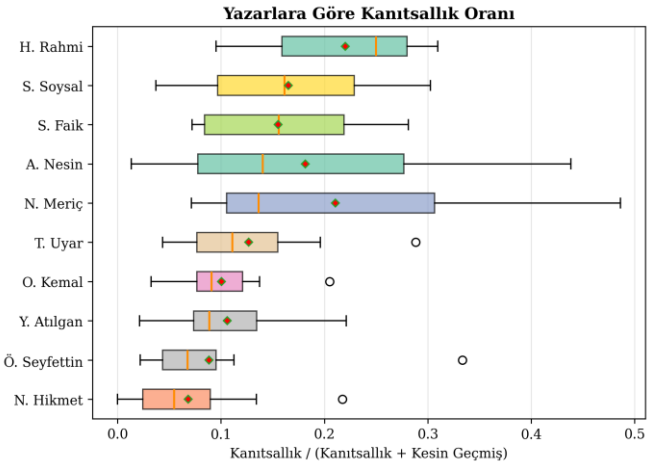
Morfolojik alt-grup analizi 47 Stanza özelliğinin beş alt gruba ayrılmasıyla gerçekleştirilmiştir. Alt-grup bazında ayrıntılı sonuçlar Ek-B'de verilmektedir. Bağımlılık ilişkileri (14 özellik) en yüksek alt-grup performansını göstermiştir (F1 = 0.409), bunu POS dağılımı (0.372), durum ekleri (0.286), zaman/görünüş/kanıtsallık (0.228) ve Türkçeye özgü özellikler (0.211) izlemiştir. Hiçbir alt-grup tek başına güçlü olmasa da birleşik morfolojik set (0.577) parçaların toplamını aşan bir performans sergilemiştir.

Şekil 2'de sunulan kanıtsallık oranı, yazarlar arasında belirgin farklılıklar sergilemektedir. Bu bulgular Türkçedeki -mIş/-DI karşıtlığına ilişkin dilbilimsel çerçeveyle tutarlıdır: -mIş eki duyum, çıkarım, dolaylı bilgiyi, başka bir deyişle dolaylılığı dilbilgisel olarak işaretleyen belirtili biçimdir [18]. -DI eki ise bu karşıtlığın belirtisiz üyesi olarak dolaylılığı olumsuzlayan, nötr bir geçmiş zaman biçimi işlevi görür [17]. Dolayısıyla yazarların -mIş/-DI tercih örüntüsü anlatıcının bilgi kaynağına ilişkin dilbilgisel tutumunu yansıtan bir üslup değişkeni oluşturmaktadır. Bu tür morfolojik tercihler karakter n-gramlarının yakalayamayacağı dilbilimsel içgörüler sunmakta ve Rudin'in [39] post-hoc açıklama yöntemleri yerine baştan yorumlanabilir özelliklerin tercih edilmesi gerektiği argümanıyla uyum göstermektedir.

İki morfolojik çözümleyicinin araç-bağımlılığını ölçmek için, paylaşılan bir token örnekleminde (2303 benzersiz token) Stanza ve Zeyrek çıktıları karşılaştırılmıştır. Zeyrek tokenların %100'ünü çözümlemiş; iki sistem lemmada %52.6 oranında uyuşmuş, %47.4 oranında uyuşmamıştır. Bu yüksek uyuşmazlığa karşın iki çözümleyici benzer ve sınırlı sınıflandırma başarısına ulaşmıştır; bu da morfolojik özelliklerin sınırının ağırlıkla araç hatasından değil, özellik düzeyinden kaynaklandığını düşündürmektedir.

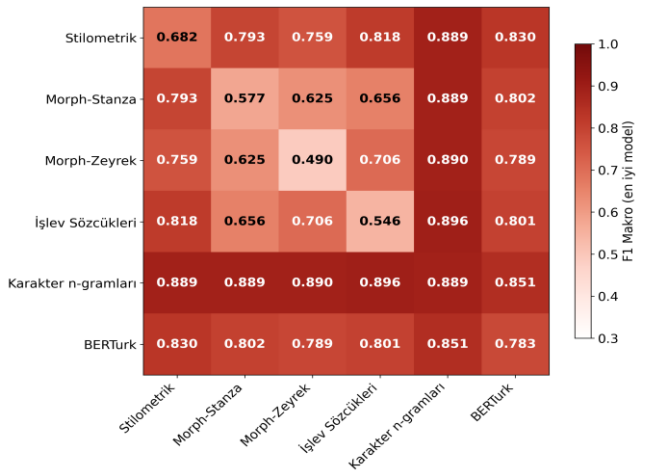
E. İkili kombinasyon analizi (Pairwise combination analysis)

İkili kombinasyon analizi, transformer (BERTurk) temsilleri ile geleneksel stilometrik/karakter-tabanlı özelliklerin birleştirilmesini, yani özellik füzyonunu da kapsamaktadır. Kombinasyon sonuçları Şekil 3'te sunulmaktadır. Karakter n-gramları herhangi bir diğer setle eşleştirildiğinde sürekli olarak 0.889–0.896 aralığında performans elde edilmiştir; bu değerler n-gramların tekil performansına çok yakındır ve diğer setlerin marjinal katkısının sınırlı olduğuna işaret etmektedir.



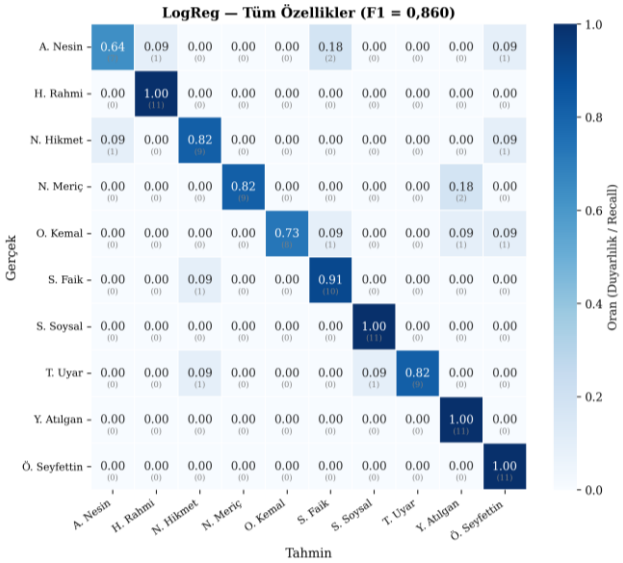
Şekil 2. Yazarlara göre kanıtsallık oranı dağılımı.

Figure 2. *Evidential ratio distribution by author.*



Şekil 3. İkili özellik seti kombinasyonları ısı haritası (en iyi model F1 makro).

Figure 3. *Pairwise feature set combination heatmap (best model F1 macro).*



Şekil 4. Karışıklık matrisi (Tüm özellikler, LogReg, 5 katlı CV toplamı).

Figure 4. *Confusion matrix (All features, LogReg, 5-fold CV aggregate).*

Morfolojik setlerin kendi aralarındaki kombinasyonu (Stanza + Zeyrek: 0.625) her iki setin tekil performansını (0.577 ve 0.491) aşmış olup; iki çözümleyicinin kısmen tamamlayıcı bilgi taşıdığını göstermektedir. Stilometrik + İşlev Sözcükleri kombinasyonu (0.818) her iki setin tekil performansını belirgin şekilde aşmıştır.

F. Sağlamlık sınaması (Robustness checks)

İki sağlamlık sınaması gerçekleştirilmiştir (Tablo 4). İlk olarak, 500 sözcüğün altındaki öyküler çıkarılarak deneyler tekrarlanmıştır (110 öykünün 95'i korunmuş; yazar başına 7–11 öykü, dağılım dengelidir). Karakter n-gramları bu alt kümede F1 = 0.847'ye düşmüştür; bu düşüş azalan örneklem boyutuna atfedilebilir.

İkinci olarak, tüm öyküler ilk 2.000 sözcükte sınırlandırılarak deneyler tekrarlanmıştır. Karakter n-gramları sınırlandırılmış metinlerle F1 = 0.907'ye yükselmiştir; bu artış, çok uzun öykülerde konu sinyalinin baskınlaşmasının sınırlandırma ile azaltılmasıyla açıklanabilir. Diğer özellik setleri de sınırlandırma ile benzer veya daha iyi sonuçlar üretmiştir. Bu sonuçlar öyküler arası sözcük sayısı dengesizliğinin sonuçları sistematik olarak bozmadığını ortaya koymaktadır.

G. Yazar bazlı performans (Per-author performance)

Şekil 4'te sunulan karışıklık matrisi birleşik modelin (LogReg) yazar bazlı ayırt edici örüntülerini yansıtmaktadır. H. Rahmi Gürpınar, Sevgi Soysal, Yusuf Atılgan ve Ömer Seyfettin tüm öykülerinin doğru sınıflandırıldığı yazarlardır (recall = 1.00). Çapraz doğrulama fold ortalamasına göre H. Rahmi ve Sevgi Soysal en yüksek yazar bazlı F1 değerlerine ulaşmıştır (0.960 ± 0.080). Diğer yüksek recall'lı yazarların (Yusuf Atılgan, Ömer Seyfettin) F1 değerleri daha düşük kalmıştır; bu durum, başka yazarların öykülerinin bu yazarlara yanlış atanmasından kaynaklanan düşük kesinlik (precision) ile açıklanmaktadır. H. Rahmi ve Sevgi Soysal'ın yüksek performansı, H. Rahmi'nin Osmanlıca söz varlığı ve kendine özgü sentaktik yapısıyla, Sevgi Soysal'ın ise deneysel anlatım üslubuyla ilişkili olabilir; ancak bu yorumlar hipotez niteliğindedir ve konu karışmasının etkisi dışarıda tutulamaz. Öte yandan en düşük performans Aziz Nesin'de gözlenmiştir (F1 = 0.660 ± 0.377); matristeki dağılıma göre Nesin'in öyküleri H. Rahmi, S. Faik ve Ö. Seyfettin ile karışmaktadır. Bu durum Nesin'in sade ve gündelik konuşma diline yakın söz varlığının diğer yazarlarla örtüşmesinden kaynaklanıyor olabilir.

H. Yazar-bazlı ayırt edici özellikler

Karışıklık matrisi yorumlarını desteklemek için her yazara ait en ayırt edici karakter n-gramları (çok-sınıflı Lojistik Regresyon katsayıları) çıkarılmıştır. Bulgular yazar üsluplarıyla tutarlıdır: Hüseyin Rahmi Gürpınar için diyalog tireleri (": – ") ve "hanım" hitabı; Yusuf Atılgan için noktalı virgül ("; ") ve yinelenen "kork-" kökü; Ömer Seyfettin için ünlem-tırnak dizileri ("!"). Bu nicel göstergeler, daha önce hipotez olarak sunulan yazar-bazlı yorumları desteklemektedir; ancak konu karışması tümüyle dışlanamaz.

İ. Aynı dönem kontrol derlemi

Yazar ayırt ediciliğinin dönem/üslup benzerliğine duyarlılığını sınamak için, erken–orta 20. yüzyıl gerçekçi kısa öykü kuşağından dört yazardan (Sabahattin Ali, Ahmet Hamdi Tanpınar, Sait Faik Abasıyanık, Ömer Seyfettin; yazar başına 11 öykü, toplam 44 öykü, ~67000 sözcük) ikinci bir kontrol derlemi oluşturulmuş ve tüm hat bu derlemde tekrarlanmıştır. Bu kohortta da karakter n-gramları en yüksek tekil başarıyı vermiştir (F1 = 0.933), bunu birleşik model (0.917) ve BERTurk (0.830) izlemiştir (Tablo 5). Ayırt ediciliğin, dönem ve tür sabitlendiğinde dahi yüksek kalması, yakalanan sinyalin büyük ölçüde bireysel üsluptan kaynaklandığını desteklemektedir. Sınıf sayısı farkı (4'e karşı 10) nedeniyle mutlak F1 değerleri doğrudan kıyaslanamaz; karşılaştırma "ayırt ediciliğin yüksek kalması" düzeyinde yorumlanmıştır.

J. Sınırlılıklar (Limitations)

Bu çalışmanın başlıca sınırlılıkları aşağıdaki gibi özetlenebilir:

Tablo 4. Sağlamlık sınaması: ≥500 sözcük ve 2.000 sözcüğe sınırlandırma.

Table 4. *Robustness checks: ≥500 words and 2,000-word truncation.*

Özellik Seti	Ana (F1)	≥500 sözcük (F1)	Sınırlı* (F1)
K. n-gramları	0.889	0.847	0.907
BERTurk	0.783	0.722	0.795
Stilometrik	0.682	0.654	0.696
Morph-Stanza	0.577	0.527	0.558
İşlev Sözcükleri	0.546	0.547	0.558
Morph-Zeyrek	0.491	0.453	0.502

K. n-gramları: Karakter n-gramları, * İlk 2000 sözcük

Tablo 5. Aynı-dönem kontrol derlemi (en iyi model F1).

Table 5. *Same-period control corpus (best model F1).*

Model / Özellik seti	F1
Karakter n-gramları	0.933
Tümü (birleşik)	0.917
İşlev Sözcükleri	0.843
BERTurk	0.830
Morph-Stanza	0.755
Stylometric	0.698
Morph-Zeyrek	0.698

- Derlem boyutu görece küçüktür (110 öykü, 10 yazar).
- Öyküler arası sözcük sayısı dengesizliği (197–11266 sözcük) mevcuttur; sağlamlık sınaması bu durumun sonuçları sistematik olarak etkilemediğini gösterse de daha dengeli bir derlem tercih edilebilir.
- Stanza UD hattının morfolojik etiketleme doğruluğu %100 değildir; belirsizlik giderme hataları özellik kalitesini olumsuz etkileyebilir [34].
- Yazar-konu karışması kontrol edilmemiştir; farklı yazarların farklı konularda yazması sınıflandırıcının konu sinyalini yakalamasına yol açabilir [1].
- n-gram morfolojik örtüşme analizi otomatik karakter dizisi eşleştirme ile gerçekleştirilmiş olup kısa n-gramların çok anlamlılığı nedeniyle kategorileme belirsizliği içermektedir.
- 5 katlı CV ile eşlenmiş bootstrap testinin istatistiksel gücü sınırlıdır; daha büyük derlemlerle tekrar değerlendirme gereklidir.
- PCA ile boyut-eşitlemeli ablasyonun sonuçları sınıflandırıcı seçimine bağlıdır. Bu çerçevenin farklı sınıflandırıcılarla tutarlılığı daha geniş veri setlerinde doğrulanmalıdır.

IV. SONUÇ (CONCLUSION)

Bu çalışmada, Türkçe kısa öykülerde yazar tanıma amacıyla altı farklı özellik seti ve beş sınıflandırıcı model sistematik olarak karşılaştırılmıştır. Elde edilen temel bulgular aşağıdaki gibi özetlenebilir:

- Karakter 2–4 gram TF-IDF, Türkçe yazar tanımada en güçlü tekil özellik setidir (F1 = 0.889). Bu sonuç karakter n-gramlarının İngilizce üzerindeki üstünlüğünün Türkçe gibi bitişken bir dilde de korunduğunu göstermektedir.
- Standart ablasyonda morfolojik, stilometrik ve işlev sözcüğü özellik setlerinin etkisiz (redundant) görünmesi, kısmen özellik setleri arasındaki boyut dengesizliğiyle açıklanabilir. PCA ile tüm setler eşit boyuta indirildiğinde, RF sınıflandırıcısı ile karakter n-gramları (ΔF1 = −0.057) ve Morph-Zeyrek (ΔF1 = −0.034) setlerinin katkısı istatistiksel olarak anlamlı bulunmuştur. Ancak bu etki sınıflandırıcıya bağlıdır. LogReg ile PCA ablasyonu standart ablasyonla benzer örüntüler sergilemiştir. Bu bulgu ablasyon çalışmalarında hem boyut dengesi hem sınıflandırıcı seçiminin göz önünde bulundurulması gerektiğini vurgulayan metodolojik bir çıkarımdır.
- En ayırt edici karakter n-gramlarının %16'sı Türkçe morfolojik ek dizileriyle ilişkiliyken %47'si noktalama örüntülerine karşılık gelmektedir. Bu sonuç n-gramların gücünün yalnızca morfolojik bilgiyi yakalamakla sınırlı olmadığını ve noktalama tercihlerinin de güçlü yazar sinyalleri taşıdığını ortaya koymaktadır.
- İki farklı morfolojik çözümleyici (Stanza UD ve Zeyrek) ile elde edilen sonuçlar tutarlıdır ve her ikisi de tekil olarak orta düzeyde sınıflandırma performansına ulaşmıştır (Stanza: 0.577, Zeyrek: 0.491). Morfolojik özellikler sınıflandırma performansı açısından sınırlı kalsa da kanıtsallık oranı ve zarf-fiil dağılımları gibi ölçümler aracılığıyla yazarlar arasındaki üslup farklarının dilbilimsel olarak yorumlanmasına olanak tanımaktadır.
- LogReg ve SVM-Doğrusal, yüksek boyutlu özellik setlerinde (karakter n-gramları, BERTurk) en iyi sonuçları vermiştir. RF ise düşük boyutlu setlerde (stilometrik, morfolojik, işlev sözcükleri) daha tutarlı performans üretmiştir. Sınıflandırıcı seçimi özellik seti boyutuyla birlikte değerlendirilmelidir.

Gelecek çalışmalarda derlem boyutunun artırılması, yazar-konu karışmasının kontrol edilmesi (konu maskeleme yöntemleriyle), karakter n-gramlarının ayırt edici gücünün kaynağını anlamak için SHAP/LIME gibi post-hoc yöntemlerin yanı sıra baştan yorumlanabilir alt-kategori analizlerinin de karşılaştırmalı olarak uygulanması, anlamsal temsillerin (BERTurk/cümle-dönüştürücü) ince ayarlı (fine-tuned) veya stilometrik özelliklerle hibritlenmiş biçimde kullanılması ve PCA ile boyut-eşitlemeli ablasyon çerçevesinin farklı diller ve metin türleri üzerinde sınanması planlanmaktadır.

YAZAR BEYANI (AUTHOR STATEMENT)

İntihal Kontrolü (Plagiarism Check)—Makale iThenticate programı ile taranmış ve derginin intihal politikası ile uyumlu bulunmuştur.
Çıkar Çatışması (Conflict of Interest)—Makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır.
Etik Kurul Onayı (Ethics Committee Approval)—Makalede etik kurul izni alınmasına gerek yoktur.
Yapay Zekâ Araçlarının Kullanımı (Use of Artificial Intelligence Tools)—Bu çalışmada dilbilgisi ve yazım denetimi için Claude (Anthropic) modeli kullanılmıştır. Tüm içerik yazar(lar)ın özgün katkısını yansıtmaktadır.
Finansman ve Proje Desteği (Funding)—Bu çalışma için herhangi bir kurumsal/finansal destek alınmamıştır.
Veri Paylaşım (Data availability)—Bu çalışmanın bulgularını destekleyen veriler yazar tarafından üretilmiş olup; makul talepler doğrultusunda yazardan temin edilebilir.
Yazar Katkısı (CRediT Author Contribution)—Kavramsallaştırma, Yöntem, Yazılım, Doğrulama, Biçimsel Analiz, Araştırma, Analiz Araçlarının Sağlanması, Veri Düzenleme, Yazma – ilk taslak, Yazma – inceleme ve düzeltme, Görselleştirme süreçleri makalenin tek yazarı tarafından gerçekleştirilmiştir.
Teşekkür (Acknowledgment)—Yazar, yapıcı yorumları ve makaleye yaptıkları iyileştirmeler için isimsiz hakemlere teşekkür eder.

KAYNAKLAR (REFERENCES)

[1] E. Stamatatos, "A survey of modern authorship attribution methods", *Journal of the American Society for Information Science and Technology*, 60(3), 538–556, 2009. https://doi.org/10.1002/asi.21001.

[2] M. Koppel, J. Schler, S. Argamon, "Computational methods in authorship attribution", *Journal of the American Society for Information Science and Technology*, 60(1), 9–26, 2009. https://doi.org/10.1002/asi.20961.

[3] X. He, A. H. Lashkari, N. Vombatkere, D. P. Sharma, "Authorship attribution methods, challenges, and future research directions: A comprehensive survey", *Information*, 15(3), 131, 2024. https://doi.org/10.3390/info15030131.

[4] B. Huang, C. Chen, K. Shu, "Authorship attribution in the era of LLMs: Problems, methodologies, and challenges", *arXiv*, 2024, https://arxiv.org/abs/2408.08946.

[5] J. Burrows, "'Delta': a measure of stylistic difference and a guide to likely authorship", *Literary and Linguistic Computing*, 17(3), 267–287, 2002. https://doi.org/10.1093/llc/17.3.267.

[6] M. Kestemont, "Function words in authorship attribution: From black magic to theory?", *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, Gothenburg, Sweden, April 2014, 59–66. https://doi.org/10.3115/v1/w14-0908.

[7] U. Sapkota, S. Bethard, M. Montes, T. Solorio, "Not all character n-grams are created equal: A study in authorship attribution", *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language Technologies (NAACL)*, Denver, USA, May-June 2015, 93–102. https://doi.org/10.3115/v1/n15-1010.

[8] E. Stamatatos, "On the robustness of authorship attribution based on character n-gram features", *Journal of Law and Policy*, 21(2), 421–439, 2013.

[9] M. Fabien, E. Villatoro-Tello, P. Motlicek, S. Parida, "BertAA: BERT fine-tuning for authorship attribution", *Proceedings of the 17th International Conference on Natural Language Processing (ICON)*, Patna, India, December 2020, 127–137. https://aclanthology.org/2020.icon-main.16/.

[10] C. Schöch, J. Dudar, E. Fileva, A. Šeļa, "Multilingual Stylometry: The Influence of Language on the Performance of Authorship Attribution using Corpora from the European Literary Text Collection (ELTeC)", *Proceedings of the Computational Humanities Research Conference 2024 (CHR 2024)*, Aarhus, Denmark, December 2024, 386-408, https://ceur-ws.org/Vol-3834/paper9.pdf.

[11] J. Rybicki, M. Eder, "Deeper Delta across genres and languages: Do we really need the most frequent words?", *Literary and Linguistic Computing*, 26(3), 315–321, 2011. https://doi.org/10.1093/llc/fqr031.

[12] S. Evert, T. Proisl, F. Jannidis, I. Reger, S. Pielström, C. Schöch, T. Vitt, "Understanding and explaining Delta measures for authorship attribution", *Digital Scholarship in the Humanities*, 32(suppl_2), ii4-ii16, 2017. https://doi.org/10.1093/llc/fqx023.

[13] K. Oflazer, "Turkish and its challenges for language processing", *Language Resources and Evaluation*, 48(4), 639–653, 2014. https://doi.org/10.1007/s10579-014-9267-2.

[14] Ç. Çöltekin, "A freely available morphological analyzer for Turkish", *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta, May 2010, 820-827. https://doi.org/10.63317/3sqjjmc7u7w5.

[15] K. Oflazer, M. Saraçlar, *Turkish Natural Language Processing*, Cham, Switzerland, Springer, 2018.

[16] H. Sak, T. Güngör, M. Saraçlar, "Turkish language resources: Morphological parser, morphological disambiguator and web corpus", *Proceedings of the 6th International Conference on Natural Language Processing (GoTAL 2008)*, Gothenburg, Sweden, 25-27 August 2008, 417-427. https://doi.org/10.1007/978-3-540-85287-2_40.

[17] L. Johanson, "Turkic indirectives", *Evidentials: Turkic, Iranian and Neighbouring Languages*, L. Johanson, B. Utas, Eds., Berlin, Germany, Mouton de Gruyter, 2000, 61-88.

[18] A. Göksel, C. Kerslake, *Turkish: A Comprehensive Grammar*, London, UK, Routledge, 2005.

[19] M. Haspelmath, E. König, *Converbs in Cross-Linguistic Perspective: Structure and Meaning of Adverbial Verb Forms*, Berlin, Germany, Mouton de Gruyter, 1995.

[20] R. Gorman, "Author identification of short texts using dependency treebanks without vocabulary", *Digital Scholarship in the Humanities*, 35(4), 812–825, 2020. https://doi.org/10.1093/llc/fqz070.

[21] M. F. Amasyalı, B. Diri, "Automatic Turkish text categorization in terms of author, genre and gender", *Proceedings of the 11th International Conference on Applications of Natural Language to Information Systems*, Klagenfurt, Austria, 31 May -02 June 2006, 221-226. https://doi.org/10.1007/11765448_22.

[22] F. Türkoğlu, B. Diri, M. F. Amasyalı, "Author attribution of Turkish texts by feature mining", *Proceedings of the 3rd International Conference on Intelligent Computing (ICIC 2007)*, Qingdao, China, 21-24 August 2007, 1086–1093. https://doi.org/10.1007/978-3-540-74171-8_110.

[23] Y. Bay, E. Çelebi, "Feature selection for enhanced author identification of Turkish text", *Proceeding of the 30th International Symposium on Computer and Information Sciences (ISCIS)*, London, England, 21-24 September 2015, 371-379. https://doi.org/10.1007/978-3-319-22635-4_34.

[24] M. Güllü, H. Polat, "Text authorship identification based on ensemble learning and genetic algorithm combination in Turkish text", *Politeknik Dergisi*, 25(3), 1287–1297, 2022. https://doi.org/10.2339/politeknik.992493.

[25] A. E. Pirli, "POS-tags, lemmatization, and feature frequencies: Stylometric analyses in Turkish using stylo", *International Journal of Digital Humanities*, 6(3), 313–334, 2025. https://doi.org/10.1007/s42803-024-00094-1.

[26] N. Ş. Saygılı, T. Acarman, T. Amghar, B. Levrat, "Taking Advantage of Turkish Characteristic Features to Tackle with Authorship Attribution Problems for Turkish", *Proceedings of ICCGI 2016: The Eleventh International Multi-Conference on Computing in the Global Information Technology*, Barcelona, Spain, November 2016, 25–29.

[27] H. Gündüz, "Scalable gender profiling from Turkish texts using deep embeddings and meta-heuristic feature selection", *Journal of Theoretical and Applied Electronic Commerce Research*, 20(4), 253, 2025. https://doi.org/10.3390/jtaer20040253.

[28] S. Yıldırım, T. Yıldız, "Türkçe için karşılaştırmalı metin sınıflandırma analizi", *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 24(5), 879-886, 2018. https://doi.org/10.5505/pajes.2018.15931.

[29] F. Can, J. M. Patton, "Change of word characteristics in 20th-century Turkish literature: A statistical analysis", *Journal of Quantitative Linguistics*, 17(3), 167–190, 2010. https://doi.org/10.1080/09296174.2010.485444.

[30] K. Altıntaş, F. Can, J. M. Patton, "Language change quantification using time-separated parallel translations", *Literary and Linguistic Computing*, 22(4), 375–393, 2007. https://doi.org/10.1093/llc/fqm026.

[31] M. Eder, "Does size matter? Authorship attribution, small samples, big problem", *Digital Scholarship in the Humanities*, 30(2), 167–182, 2015. https://doi.org/10.1093/llc/fqt066.

[32] K. Kettunen, "Can type-token ratio be used to show morphological complexity of languages?", *Journal of Quantitative Linguistics*, 21(3), 223–245, 2014. https://doi.org/10.1080/09296174.2014.911506.

[33] G. U. Yule, *The Statistical Study of Literary Vocabulary*, Cambridge, UK, Cambridge University Press, 1944.

[34] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, "Stanza: A Python natural language processing toolkit for many human languages", *Proceedings of the 58th annual meeting of the association for computational linguistics: system demonstrations*, Online, July 2020, 101–108. https://doi.org/10.18653/v1/2020.acl-demos.14.

[35] O. Bulat, "Zeyrek: Morphological analyzer and lemmatizer for Turkish (Python port of Zemberek-NLP)", GitHub repository, https://github.com/obulat/zeyrek (26.04.2026).

[36] A. A. Akın, M. D. Akın, "Zemberek, an open source NLP framework for Turkic languages", https://github.com/ahmetaa/zemberek-nlp.

[37] S. Schweter, "BERTurk – BERT models for Turkish", *Zenodo*, 2020, https://doi.org/10.5281/zenodo.3770924.

[38] E. Çelik, "bert-base-turkish-cased-mean-nli-stsb-tr: Türkçe cümle-dönüştürücü (sentence-transformers) modeli", *Hugging Face Model Hub*, https://huggingface.co/emrecan/bert-base-turkish-cased-mean-nli-stsb-tr (23.06.2026).

[39] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", *Nature Machine Intelligence*, 1(5), 206–215, 2019. https://doi.org/10.1038/s42256-019-0048-x.

EK-A

Tablo A1. Tüm özellik seti × sınıflandırıcı kombinasyonları (5 katlı CV).

Table A1. *All feature set × classifier combinations (5-fold stratified CV).*

Özellik Seti	Model	F1 Makro	Doğruluk
Stilometrik	SVM-Linear	0.596 ± 0.078	0.618 ± 0.074
Stilometrik	SVM-RBF	0.582 ± 0.109	0.591 ± 0.111
Stilometrik	LogReg	0.623 ± 0.059	0.645 ± 0.053
Stilometrik	RF	**0.682 ± 0.050**	**0.700 ± 0.055**
Stilometrik	MLP	0.520 ± 0.054	0.564 ± 0.062
Morph-Stanza	SVM-Linear	0.532 ± 0.080	0.573 ± 0.074
Morph-Stanza	SVM-RBF	0.518 ± 0.075	0.555 ± 0.088
Morph-Stanza	LogReg	0.476 ± 0.076	0.509 ± 0.078
Morph-Stanza	RF	**0.577 ± 0.123**	**0.591 ± 0.111**
Morph-Stanza	MLP	0.365 ± 0.143	0.409 ± 0.135
Morph-Zeyrek	SVM-Linear	0.454 ± 0.124	0.464 ± 0.123
Morph-Zeyrek	SVM-RBF	0.474 ± 0.049	0.473 ± 0.036
Morph-Zeyrek	LogReg	0.471 ± 0.119	0.491 ± 0.113
Morph-Zeyrek	RF	**0.491 ± 0.136**	**0.518 ± 0.130**
Morph-Zeyrek	MLP	0.298 ± 0.097	0.345 ± 0.102
İşlev Sözcükleri	SVM-Linear	0.533 ± 0.093	0.573 ± 0.084
İşlev Sözcükleri	SVM-RBF	0.452 ± 0.081	0.491 ± 0.078
İşlev Sözcükleri	LogReg	0.520 ± 0.094	0.555 ± 0.088
İşlev Sözcükleri	RF	**0.546 ± 0.070**	**0.582 ± 0.073**
İşlev Sözcükleri	MLP	0.383 ± 0.108	0.427 ± 0.124
K. n-gramları	SVM-Linear	0.851 ± 0.071	0.855 ± 0.067
K. n-gramları	SVM-RBF	0.710 ± 0.055	0.727 ± 0.029
K. n-gramları	LogReg	**0.889 ± 0.049**	**0.900 ± 0.034**
K. n-gramları	RF	0.809 ± 0.091	0.827 ± 0.078
K. n-gramları	MLP	0.686 ± 0.123	0.700 ± 0.110
BERTurk	SVM-Linear	**0.783 ± 0.081**	**0.791 ± 0.074**
BERTurk	SVM-RBF	0.716 ± 0.075	0.718 ± 0.053
BERTurk	LogReg	0.777 ± 0.068	0.782 ± 0.067
BERTurk	RF	0.686 ± 0.094	0.700 ± 0.074
BERTurk	MLP	0.567 ± 0.073	0.591 ± 0.041
Tüm Özellikler	SVM-Linear	0.830 ± 0.066	0.836 ± 0.062
Tüm Özellikler	SVM-RBF	0.747 ± 0.069	0.764 ± 0.060
Tüm Özellikler	LogReg	**0.860 ± 0.052**	**0.873 ± 0.045**
Tüm Özellikler	RF	0.780 ± 0.062	0.800 ± 0.036
Tüm Özellikler	MLP	0.678 ± 0.137	0.709 ± 0.117

Tablo A2. Morfolojik özellik alt-grup analizi (Stanza UD, en iyi model).

Table A2. *Morphological feature subgroup analysis (Stanza UD, best model).*

Alt Grup	En İyi Model	Özellik Sayısı	F1 Makro
POS Dağılımı	RF	14	0.372 ± 0.099
Bağımlılık İlişkileri	LogReg	14	0.409 ± 0.042
Zaman/Görünüş/Kanıtsallık	SVM-RBF	9	0.228 ± 0.081
Durum Ekleri	RF	7	0.286 ± 0.049
Türkçeye Özgü	SVM-Linear	3	0.211 ± 0.057
Tüm Morfolojik: Stanza	RF	47	0.577 ± 0.123

Tablo A3. Veri seti kaynak ve baskı künyeleri.

Table A3. *Dataset sources and editions.*

Yazar	Kaynak Eserler ve Ayrıntıları
Sait Faik Abasıyanık	Mahalle Kahvesi (1. basım, 2012; 978-625-405-404-4); Alemdağ'da Var Bir Yılan (978-605-360-720-5); Seçme Hikâyeler (978-605-360-724-3); Türkiye İş Bankası Kültür Yayınları
Hüseyin Rahmi Gürpınar	Hikâyeler (Ankara, 2022; 978-975-17-5363-2); Türk Dil Kurumu Yayınları
Tomris Uyar	Bütün Öyküleri (1. baskı, İstanbul, 2014; 978-975-08-3074-7); Yapı Kredi Yayınları
Aziz Nesin	Memleketin Birinde (978-605-4702-43-5); Sizin Memlekette Eşek Yok mu? (978-605-4702-13-8); Nesin Yayınevi
Ömer Seyfettin	Seçme Hikâyeler (978-605-332-214-6); Türkiye İş Bankası Kültür Yayınları
Nâzım Hikmet	Hikâyeler (3. basım, 1993; 975-418-072-5); Anadolu Yayıncılık
Nezihe Meriç	Toplu Öyküleri I–II (1. baskı, İstanbul, 1998; 975-363-805-1); Yapı Kredi Yayınları
Sevgi Soysal	Tante Rosa (2002; 978-975-05-0092-3); Barış Adlı Çocuk (2003; 978-975-05-0197-5); İletişim Yayınları
Orhan Kemal	Çamaşırcının Kızı (5. basım, İstanbul, 1998; 975-478-102-8); Tekin Yayınevi
Yusuf Atılgan	Bütün Öyküleri (İstanbul, 2017; 978-975-363-790-9); Yapı Kredi Yayınları