IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 March 2025, accepted 29 March 2025, date of publication 1 April 2025, date of current version 10 April 2025.
Digital Object Identifier 10.1109/ACCESS.2025.3556725

RESEARCH ARTICLE

Summarization, Prediction, and Analysis of Turkish Constitutional Court Decisions With Explainable Artificial Intelligence and a Hybrid Natural Language Processing Method

TÜLAY TURAN [1] AND ECİR UĞUR KÜÇÜKSİLLE [2]
[1]Department of Computer Engineering, Burdur Mehmet Akif Ersoy University, 15200 Burdur, Türkiye
[2]Department of Computer Engineering, Suleyman Demirel University, 32260 Isparta, Türkiye

Corresponding author: Tülay Turan (tulayturan@mehmetakif.edu.tr)

This study is derived from Tülay TURAN's doctoral thesis titled 'Legal Text Analysis with Explainable Artificial Intelligence' in the Computer Engineering program at Süleyman Demirel University.

ABSTRACT The use of artificial intelligence in legal analysis represents a significant transformation process in modern legal practices. The literature review showed that the number of studies on Turkish court decisions was limited, and only classification applications were developed. This study aims to summarize, predict, and analyze the Turkish Constitutional Court's decision texts. In this context, first a new, unique dataset was created. The dataset included class labels, original court decision texts, and summary court decision texts created by expert lawyers. Then, a hybrid summary model was established, combining extractive and abstract summarization to summarize long court decisions automatically. With this model, the summarization efficiency was increased, and the token limitation problem common in existing transformative models was eliminated. The model showed good performance for summarization by obtaining Rouge-1, Rouge-2, and Rouge-L scores of 0.6129, 0.5884, and 0.5891 respectively. After the summarization phase, the Legal Judgment Prediction applications were developed. Separate classification models were developed for court decision texts and summary court decision texts, obtained using the hybrid model. When the models were compared, the XGBoost model achieved the best performance in legal judgment prediction tasks, with an accuracy rate of 93.84% for full texts and 62.30% for summary texts. In the final stage of the study, the model results were explained using the SHapley Additive exPlanations method. The findings of this study emphasize the superiority of hybrid approaches to legal document analysis. They highlighted the vital role of explainable techniques in improving transparency and reliability in legal processes.

INDEX TERMS Explainable artificial intelligence, hybrid document summarization, legal judgment prediction, natural language processing, Turkish constitutional court.

I. INTRODUCTION

Virtual legal analysis has started a transformative period because of improvements in natural language processing (NLP) and expanded legal data accessibility throughout the world [1]. The technology advancement has created a fundamental change which enables legal operators to complete their work more quickly and effectively with reduced financial burden. The use of artificial intelligence (AI) applications continues to grow within legal domains including case decision forecasting together with document analysis and legal support and contract management systems which transform legal service abilities.

The development of Turkish legal AI solutions progresses slowly as the world observes significant progress in AI-based legal applications. A major impediment exists due to the insufficient availability of Turkish legal data collections. Current datasets serve classification needs but they do not

The associate editor coordinating the review of this manuscript and approving it for publication was Anandakumar Haldorai.

contain sufficient analytical summaries that would allow detailed legal evaluation. The complex nature and extended length as well as significant number of citations within judicial texts make legal document summarization an ongoing difficult problem. Strategic solutions need to be developed for legal documents because standard summarization models do not understand legal syntax and linguistic expressions properly. The prediction capabilities of legal judgment prediction (LJP) models face continuous improvements but continue to encounter challenges when attempting to interpret complex judicial decision-making processes. AI applications find limited adoption within legal practice because they lack sufficient explanation features in their operational logic.

Fast improvements of artificial intelligence systems to analyze law do not solve all problems related to Turkish legal AI applications. Available Turkish legal text datasets currently have no decision summaries which reduces their potential value in performing complete AI assessments. The preceding research on legal summarization concentrates on scientific papers together with news content yet omits the distinct features that make legal documents different. The conventional LJP models lack capability to execute legal reasoning properly while AI-based legal decision systems operating without explainability diminish their practical use.

To address these gaps, this study introduces a novel Turkish legal dataset enriched with decision summaries, facilitating a more nuanced AI-based legal analysis. A hybrid summarization model is proposed, combining inferential summarization with abstract refinement to improve summarization efficiency while addressing token limitations in transformer-based models. Additionally, LJP models optimized for both full legal texts and their summaries are developed, incorporating hyperparameter optimization to enhance predictive performance. Finally, explainable AI techniques using SHapley Additive exPlanations (SHAP) values are integrated, transforming black-box AI decisions into interpretable insights for legal professionals and clients.

In Section II, a detailed literature review is provided on AI applications in legal analysis, including legal document summarization, judgment prediction models, and explainability techniques such as SHAP. Section III discusses the methodology, including dataset construction, model selection, and experimental setup. Section IV. experiments presents the results and analysis, with comparative evaluations summarized in Tables 6 and 7. Section V provides an in-depth discussion, covering the implications of findings, limitations, and policy considerations. Finally, Section VI concludes with key findings and outlines future research directions, focusing on dataset expansion, additional AI model exploration, and real-world applications in legal decision support systems.

II. RELATED WORKS

The application of AI in the legal domain has gained significant traction in recent years. Various courts, including the European Court of Human Rights (ECHR) [9], [10], [11], the Supreme People's Court of China [12], [13], [14], and the Supreme Court of the United States [15], [16], [17], have enabled access to judicial decisions, facilitating AI-driven legal applications. Most AI-based legal studies have focused on English and Chinese datasets, while research on Turkish legal AI remains sparse [18], [19], [20].

Automatic Text Summarization (ATS) has been extensively studied in NLP, primarily focusing on news articles [21], [22], [23] and scientific texts [24], [25], [26]. However, legal document summarization presents unique challenges due to the length, technical language, and citation-heavy nature of legal texts. Existing summarization models, such as CaseSummarizer [27] and DELSum [28], highlight the potential of AI-driven summarization but also reveal limitations in handling extensive legal documents [29], [30].

LJP has become a critical AI application, aiming to predict court decisions based on case facts and legal precedents. Modern LJP models utilize deep learning as well as transformer-based architectures for their operational methods after initial rule-based and statistical models [31], [32], [33]. The modern AI systems demonstrate better accuracy results but they don't reproduce judicial reasoning precisely which shows why better AI techniques are necessary.

Legal professionals need AI models in their legal analysis to operate with transparency because explainability remains essential. Explainable artificial intelligence (XAI) techniques have been developed through SHAP to improve AI interpretability as discussed in [34] and [35]. SHAP values make black box AI become white box systems through showing AI predictions to legal practitioners thus increasing their trust in the system while boosting system usability [36], [37].

Recent studies have demonstrated the effectiveness of SHAP in various machine learning applications beyond the legal domain. For instance, SHAP has been successfully used in fuel heating value prediction [38], financial forecasting such as Bitcoin price prediction [39], water quality assessment [40], audit opinion prediction [41], and IoT security enhancements [42]. These studies underline the versatility of SHAP in providing model interpretability and improving decision-making transparency across multiple fields. The success of SHAP in these domains strengthens its case for application in legal AI, where explainability and interpretability are paramount. This study builds upon these advancements by integrating SHAP into legal judgment prediction, ensuring that AI-driven decisions are not only accurate but also interpretable for legal professionals.

In conclusion, the convergence of artificial intelligence techniques in the legal domain points to a future of considerable potential. With ongoing advancements, particularly in LJP and ATS, alongside the integration of explainability tools like SHAP, the field is poised for transformation. However, addressing the existing gaps, particularly in less-studied jurisdictions such as Turkey, and in the unique demands of legal text processing will be essential to capitalizing on the opportunities presented by these technological advancements.

III. MATERIALS AND METHODS

A. TEXT SUMMARIZATION WITH NATURAL LANGUAGE PROCESSING

Text summarization aims to divide long textual documents into concise, coherent, and accurate summaries. With the exponential growth of textual data, automatic text summarization has become indispensable for efficient processing and extracting valuable information [43]. There are two primary approaches to text summarization: extractive and abstractive. Extractive summarization involves selecting critical sentences from the original document to form a summary, focusing on capturing the most crucial information without altering the original text [44]. On the other hand, abstractive summarization seeks to paraphrase and reinterpret the essence of a text, akin to how humans might summarize by generating new sentences that encapsulate core ideas [45]. Figure 1 illustrates the basic workflow of extractive summarization, whereas Figure 2 depicts the process of abstractive summarization.

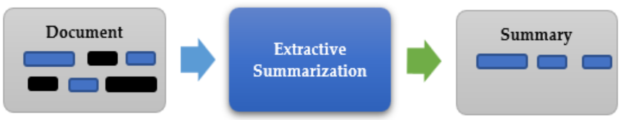


FIGURE 1. The basic workflow of extractive summarization.

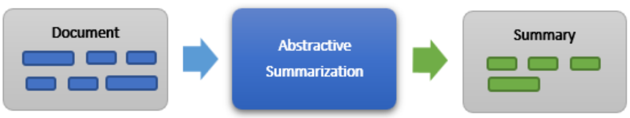


FIGURE 2. The basic workflow of abstractive summarization.

The evolution of deep learning has significantly advanced the field of text summarization, mainly through the development of sequence-to-sequence (Seq2Seq) models. Pre-trained language models (GPT, BERT, XLM) and pre-trained Seq2Seq models (BART, T5) have allowed us to overcome some of the shortcomings in neural summarization and generate higher-quality summaries [46], [47]. GPT-3/4 [48], [49], Hugging Face's Transformers [50], [51], [52], Tensorflow [53], and PyTextRank [54] have been seen as the most frequently used libraries for summarization in the literature.

In their landmark paper "All You Need is Attention", researchers from Google positioned that attention mechanisms alone could suffice for understanding sequences, introducing the transformer model architecture that leverages self-attention to process all inputs simultaneously, unlike traditional neural networks. This architecture, characterized by its efficiency and parallel processing capabilities, has set new standards in NLP applications, including text summarization [55]. In this study, the Spacy library was utilized for text preprocessing, the Bert2Bert model for generating summaries, and Rouge metrics for evaluating the fidelity and coherence of summaries relative to the original legal texts.

1) SPACY LIBRARY

Spacy is an open-source natural language processing library. It is a library containing practical text-processing tools with multilanguage support [56]. It is used in various tasks such as text preprocessing, information extraction, and text classification in natural language processing. To facilitate the analysis of the text, it enables the tasks of converting uppercase and lowercase letters, deleting punctuation marks and, numbers, and removing stop words. Spacy includes multi-task learning with pre-trained transformers, such as Bidirectional Encoder Representations from Transformers (BERT).

2) BERT2BERT

Bert2Bert is an extended version of the BERT model. BERT was developed by a Google artificial intelligence team [57]. It can be defined as a bidirectional transformer network that is pre-trained on a large corpus. Positional encoding was used to label the data elements entering and leaving the network. By tracking these labels, the attentional units calculate an algebraic map of each element's relationship with other elements. Attention queries are executed in parallel by calculating a matrix of equations called multi-head attention. Multi-headed attention ensures that different parallel calculations are performed for the same word, the results are transmitted to SoftMax, and the most appropriate result is obtained. Multilingual and monolingual models have been developed on the Bert2Bert basic architecture. The monolingual Bert2Bert models were pre-trained in the selected language only. Text summarization applications can be developed using the Bert2Bert language model, pre-trained in Turkish.

3) ROUGE METRIC

In the natural language processing domain, the quest to measure the quality of text summarization and machine translation has been significantly advanced by introducing the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metric. Conceptualized by Lin in 2004, ROUGE offers a systematic approach for assessing the closeness of machine-generated summaries to human-crafted reference summaries [58]. This metric serves as a cornerstone for evaluating NLP tasks, facilitating quantitative analysis of summarization efficacy.

At the core of the research, the ROUGE-N variant of this metric was employed as a primary tool to evaluate the quality of summarizations produced by the models. The formulation of ROUGE-N is predicated on the comparison of n-grams in machine-generated summaries against those in a corpus of reference summaries, as delineated in the following equation:

$$\mathrm{ROUGE-N} = \frac{\sum_{S \in \mathrm{Refgram}_n \in S} \sum \mathrm{Count}_{\mathrm{match}}(\mathrm{gram}_n)}{\sum_{S \in \mathrm{Refgram}_n \in S} \sum \mathrm{Count}(\mathrm{gram}_n)} \quad (1)$$

where N denotes the n-gram length under consideration, reflecting the scope of the textual units from unigrams to more extensive sequences for analysis. Variable Ref represents the collection of reference summaries against which the summarization outputs are evaluated. The numerator $Count_{match}(gram_n)$ quantifies the frequency of matching n-grams found in both the model-generated and reference summaries, thereby providing a measure of summarization recall [59].

The adoption of ROUGE-N in the study is instrumental in quantitatively gauging summarization performance, enabling an objective comparison of the informational content preserved in the generated summaries relative to the benchmarks set by human expertise. Through this metric, the goal is to elucidate how NLP models capture the salient points of legal texts, as reflected in their concordance with reference summaries.

B. CLASSIFICATION WITH NLP

Text classification within the realm of natural language processing involves fundamentally assigning predefined categories or labels to textual documents, sentences, or phrases contingent upon their inherent content. The primary objective of text classification is to facilitate automated determination of the relevant class or category for a given text, thereby streamlining the process of organizing, managing, and retrieving textual data based on thematic or semantic criteria. In this investigation, advanced classification models, including XGBoost, Multilayer Perceptron (MLP), and BERT, were leveraged to predict court decisions, showcasing their efficacy in legal judgment prediction tasks.

The choice of XGBoost, MLP, and BERT for predicting court decisions and legal decision tasks stems from their unique and complementary strengths. XGBoost is renowned for its computational efficiency and ability to handle large datasets, offering the high performance and robustness required for the complex and voluminous nature of legal texts. MLPs offer versatility and the ability to model nonlinear relationships, allowing them to capture complex patterns and nuances in legal language. With its superior contextual understanding and prior training on large-scale corpora, BERT excels in interpreting the complex and context-dependent nature of legal documents Therefore, a comparison of these models was conducted.

1) XGBOOST CLASSIFICATION ALGORITHM

Introduced by Chen and Guestrin in 2016, XGBoost is a paradigmatic advancement in machine learning, specifically designed to enhance the predictive performance of ensemble learning models by systematically incorporating weak learners, predominantly decision trees, into a robust learning framework [60].

One of the key reasons for selecting XGBoost over alternative classifiers such as AdaBoost, CatBoost, and LightGBM is its computational efficiency, scalability, and superior handling of large-scale datasets with high dimensionality. XGBoost offers several advantages, including:

- Regularization: Unlike AdaBoost, which primarily focuses on reducing bias, XGBoost integrates both L1 (Lasso) and L2 (Ridge) regularization techniques, reducing the risk of overfitting while improving model generalizability [61].
- Handling Missing Values: Unlike many other classifiers, XGBoost can efficiently process missing data and automatically learn optimal split directions, which is particularly useful for complex legal datasets.
- Parallel Processing & Speed: Compared to LightGBM, which implements a leaf-wise tree growth strategy, XGBoost's level-wise approach balances speed with accuracy, making it an optimal choice for the dataset [62].
- Tree Pruning: Unlike CatBoost, which primarily focuses on categorical data, XGBoost employs an advanced tree pruning technique that prevents overfitting by pruning nodes with minimal gain, ensuring model stability.

XGBoost has demonstrated remarkable performance in various real-world applications, including financial fraud detection [63], medical diagnosis prediction [64], and NLP-based text classification [57]. These successful implementations further justify the selection of XGBoost as a primary classifier in legal judgment prediction tasks.

2) MLP CLASSIFICATION ALGORITHM

MLP is a type of deep learning classifier that is frequently employed as the first approach baseline classifier model. MLP is a good lower order model for classification problems and has much more simplicity compared to efficient architectures such as transformers. It is also efficient to compute, able to fit non-linear functions and does reasonably well on most classification tasks.

In contrast to the more complex architectures such as BERT, MLP is less costly in terms of computation. With regards to MLP's accuracy relative to BERT's accuracy, MLP serves as a good benchmark model. With the dataset in mind, MLP proves to be more efficient in training time and accuracy, serving as a suitable base model for the metric computations with more complex discriminative classifiers.

Also, MLP models can capture several phenomena which features, representations of text are sufficiently specified. MLP does not have the advance sequential processing capability which is performed by transformers. However, MLP is an easier algorithm to understand and interpret. Hence, it aids in conducting the comparative analyses.

3) BERT CLASSIFICATION ALGORITHM

BERT (Bidirectional Encoder Representations from Transformers) was developed by Cai et al. in 2018 and has since become a state-of-the-art model for various NLP tasks, including text classification, sentiment analysis, and question answering [65]. Unlike sequential models based on LSTMs or CNNs, BERT uses a transformer model that permits context learning in both directions. The model processes the preceding and following words and phrases which form a coherent context. This feature is particularly useful in legal texts, which are usually long, vague, and depend on contexts.

The choice of BERT as opposed to other NLP models stems from its unrivaled contextual comprehension. With pretrained large-scale corpora, BERT understands complex language systems and relationships, making it excellent in predicting outcomes of legal judgments for there is a need for interpretation and context. For the same reason, BERT embeddings outperform legal documents in comparison to traditional word embeddings like Word2Vec or GloVe, which offer shallow semantic representation, since context and interpretation is everything in complex legal documents.

Moreover, BERT has been used in numerous NLP related legal research exercises, such as analyzing contracts, summarizing case law, and classifying legal documents. Because of this, BERT was incorporated as one of the classification models to leverage its advanced language comprehension skills in the field of legal AI.

4) CLASSIFICATION EVALUATION

When dealing with machine learning models, especially those utilized for classification, it is crucial to have an evaluation scorecard which helps in assessing the predictive accuracy and general effectiveness of the models [66]. For this purpose, the study utilized a scorecard containing a multitude of metrics; namely, Accuracy, Precision, Recall, and F1 score which is commonly known as the harmonic mean of precision and recall. Each of them has its own perspectives regarding the performance parameters of the classification models being analyzed.

Accuracy is one of the most important metrics, as it measures the ratio of results (irrespective of true or negative) to the processes conducted hence, offering a snapshot of the basic effectiveness in making predictions. It's measured through the equation stated below: where TP refers to true positives, TN is true negatives, FP is false positives, and FN is false negatives. It offers a straightforward measure of the overall effectiveness in making correct predictions, encapsulated by the following equation:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

Precision, also known as the positive predictive value, assesses the accuracy of the model in predicting positive labels and reflects the proportion of accurate positive predictions about the total number of positive predictions made. This metric is particularly insightful in contexts where the cost of false positives is high. Precision was calculated as follows:

$$Preccion = \frac{TP}{TP + FP} \tag{3}$$

Recall or sensitivity measures the model's capability to correctly identify all relevant instances, providing an index of the coverage of true-positive observations. This is especially crucial when missing a positive instance (false negative), which carries significant consequences. The formula for recall is given by

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

The F1-Score harmonizes the balance between precision and recall, offering a single metric that encapsulates the model's precision and ability to recall positive instances. This is particularly valuable when seeking a balance between precision and recall.

These metrics collectively provide a multifaceted view of the performance of classification models, enabling a nuanced analysis that transcends mere accuracy to consider the balance and trade-offs between different aspects of predictive performance.

C. EXPLAINABLE ARTIFICIAL INTELLIGENCE

As the complexity of AI systems increases, it is essential to understand the inner workings of these systems and ensure reliability among users, administrators, and society. Explainable artificial intelligence is the ability to explain algorithms, models, and decision processes in a way that is understandable to humans [67]. This article explains the model results obtained using the SHAP method, which is an explainable artificial intelligence techniques.

1) SHAP METHOD

According to the principles of cooperative game theory, SHAP serves as a powerful tool for unraveling the intricacies of AI models, which are often perceived as inscrutable 'black boxes.' It endeavors to enhance the transparency of these models by quantifying the influence of each feature on predictive outcomes [68]. SHAP operationalizes this by leveraging Shapley values, a concept from game theory that allocates payouts (predictions, in this context) to players (features) based on their contribution to the 'game' (prediction task).

The mathematical formulation of the Shapley value for a feature i, denoted as ϕ_{i}, is expressed as Equation 5:

$$\emptyset_i = \frac{1}{|N|!} \sum_{S \subseteq N\{i\}} |S|!\,(|N| - |S| - 1)![f\left(S \cup \left\{\mathrm{i}\right\}\right) - f(S)] \tag{5}$$

In this equation, f(S) represents the model's output for a subset of features S and N denotes the set of all features. The Shapley value, ϕi, thus signifies the mean marginal contribution of feature i across all conceivable permutations of the feature set. This calculation methodically ascertains the impact of incorporating each feature into the model, providing a nuanced understanding of how specific attributes influence the model's predictions.

IV. EXPERIMENTS

This study consisted of four main parts. A data set containing 1300 case texts and manual summaries was created in the first part. In the second part, summaries of legal decisions were obtained using the hybrid method. In the third part, describes the development of LJP models for full-text and summary text legal decisions. The results obtained with the SHAP plots in the explainable artificial intelligence techniques are explained in the fourth chapter. The steps of the study are illustrated in Figure 3.

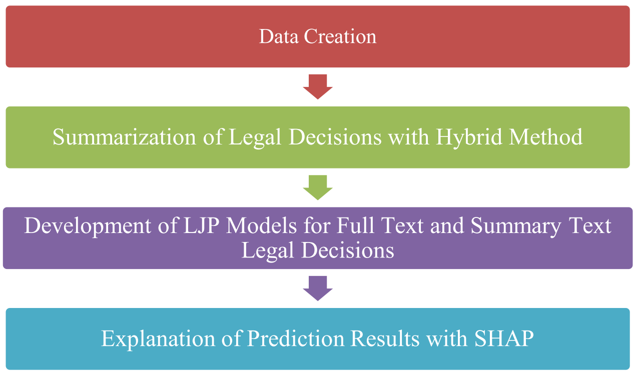


FIGURE 3. The procedural flow and methodologies employed in the study.

A. DATASET

The Turkish judicial system is divided into branches: judicial judiciary (First Instance Courts, Regional Courts of Justice, and Supreme Court of Appeals), administrative judiciary (First Instance Courts, Regional Administrative Courts, and Council of State), constitutional judiciary (Constitutional Court), and dispute judiciary (Dispute Court). While creating the study dataset, the online sharing decisions of all the courts were examined. As a result of the examinations, it was decided to use the decisions of the Constitutional Court, which stood out with their content order and number of examples. The data set was created using individual application decisions, resulting in 650 "Violations" and 650 "Non-Violations" from the Decisions Information Bank web page. The decision text consisted of six main sections. In the study, the "Examination and Justification" section was used as the "independent variable," and the "Judgment" section was used as the "dependent variable." In line with this decision, the cover page, subject of the application, application process, events and facts, and relevant legal information that is unnecessary for the study have been removed from the decision texts. To develop models for text summarization applications with artificial intelligence, summary information must be included in the data set. For this purpose, manual summaries of 1300 individual application decisions in the data set were created with the support of expert legal professionals and added to the content of the data set. Figure 4 illustrates the composition of the dataset.

Existing Turkish legal datasets are limited in scope and typically consist only of decision and class labels. This restricts their use to basic classification tasks and limits their applicability for advanced NLP applications, such as summarization, which requires more comprehensive data.

The new dataset presented in the study addresses these limitations by including not only decision texts but also expert-generated summaries of these decisions alongside class labels. This enriched dataset enables researchers to explore both classification and summarization tasks, thereby broadening the potential for advanced research and application development in the Turkish legal domain.

B. SUMMARIZATION OF LEGAL DECISIONS VIA A HYBRID METHODOLOGICAL APPROACH

Judicial decisions, characterized by their extensive length and complex content, necessitate a methodical approach to summarization that ensures that all pertinent information is succinctly captured. Traditional abstractive summarization models, even those leveraging state-of-the-art pre-trained language models, encounter limitations with input text length, notably a maximum token threshold of 512 tokens for BERT and similar transformer architectures. To address this limitation, this study introduces a hybrid model that synergistically combines inferential and abstractive summarization techniques, transcending the token input constraint. The workflow of the proposed hybrid system is illustrated in Figure 5.

In the deployed hybrid system, the summarization process is initiated with a comprehensive inferential summarization of the legal judgment texts. This methodology facilitates the aggregation of critical information dispersed throughout extensive legal decision documents. Central to this process is the employment of the first Turkish spaCy transformer model, tr_core_web_trf, released in May 2023. The significance of transformer-type tokens lies in their ability to represent lexical items, regardless of their frequency within the language, by segmenting them into subword units (WordPiece). Utilizing a WordPiece tokenization schema within the tr_core_web_trf model significantly enhances the application by accommodating a broad lexical scope. The model's training regimen was conducted on the BOUN treebank and two Named Entity Recognition (NER) corpora, WikiAnn/PanX and the Turkish Wiki NER Dataset. This comprehensive training enables the model to execute an array of NLP tasks, including part-of-the-speech (POS) tagging, dependency parsing, morphological analysis, statistical lemmatization, and NER were integrated within the pipeline. Moreover, the underpinning of the tr_core_web_trf model is the pre-trained dbmz/bert-base-turkish-cased transformer architecture, which provides a robust foundation for inferential summarization tasks. The architecture of the inferential summarization approach adopted in this study is illustrated in Figure 6, which illustrate the system's structured process and underlying components.

ozet	davametni	karar
gezi parkı eylemleri sırasında gaz kapsülü veya plastik m	64. mahkemenin 17/3/2021 tarihinde yapmış olduğu toplantıda	1
zamanında ve yeterli tıbbi tedavi alamadığından yakınma	27. mahkemenin 9/7/2020 tarihinde yapmış olduğu toplantıda	1
yargılamanın uzun sürdüğünü belirterek makul sürede y	22. mahkemenin 21/4/2021 tarihinde yapmış olduğu toplantıda	1
infaz görevlileri tarafından şiddete uğramasına rağmen a	33. mahkemenin 29/9/2020 tarihinde yapmış olduğu toplantıda	1
üyesi olmadığı ancak zaman zaman uğradığı bir derneğin	19. mahkemenin 3/6/2020 tarihinde yapmış olduğu toplantıda	1
ırak'ta ebu gureyb cezaevinde istihbarat görevlileri taraf	32. mahkemenin 8/7/2020 tarihinde yapmış olduğu toplantıda	0
mektup gönderme ve alma hakkının keyfî olarak haftada	22. mahkemenin 28/1/2020 tarihinde yapmış olduğu toplantıda	0
yapmış olduğu paylaşımla öğrencilerini ya da başka herh	22. anayasa mahkemesinin 13/1/2022 tarihinde yapmış olduğu	0
ilgili mevzuata göre avukatıyla telefon vasıtasıyla görüşn	25. mahkemenin 2/6/2020 tarihinde yapmış olduğu toplantıda	0
aynı konuya ilişkin olarak daha önce soruşturma yürütül	43. mahkemenin 15/9/2021 tarihinde yapmış olduğu toplantıda	0

FIGURE 4. Detailed content of the top 10 examples in the dataset, including case texts, summaries, and decision results (violation, non-violation).

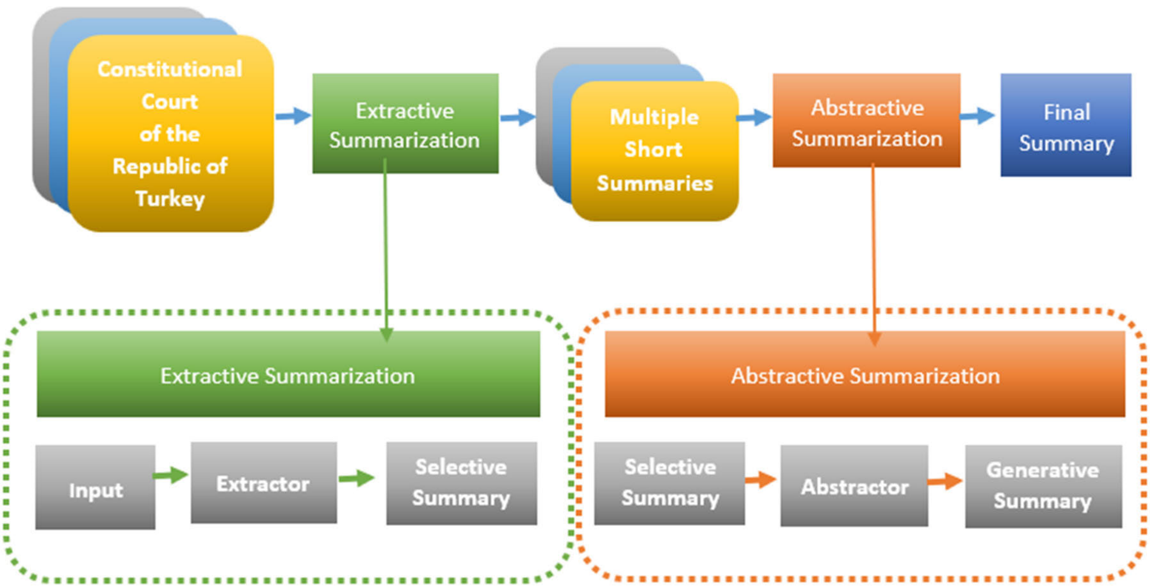


FIGURE 5. The hybrid model workflow we created to summarize the decisions of the Constitutional Court of Turkey.

A demonstrative example of the inferential summarization process is presented in Figure 7, which shows the original decision text, and Figure 8 shows the resultant inferential summary. Here, a document initially comprising 110 sentences is effectively condensed to a mere 11 sentences post-summarization. An abstract summary of the original decision text is shown in Figure 9.

In the abstraction phase of the summarization methodology, the Turkish bert2bert, a pre-trained language model renowned for its proficiency in generative tasks, was engaged. This phase builds upon the foundation laid by the inferential summarization module, which provides the initial condensed summaries as input. To effectively prime the bert2bert model for the summarization task, a task-specific prefix was affixed to the input data, facilitating the model's recognition of the desired operation. This procedural step is critical, as the model possesses the versatility to perform many NLP tasks, contingent upon the receipt of appropriate task prompts. The bert2bert model was fine-tuned using the

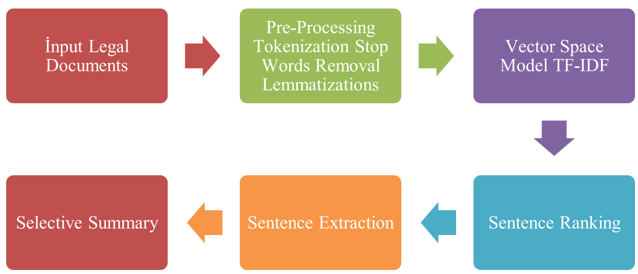
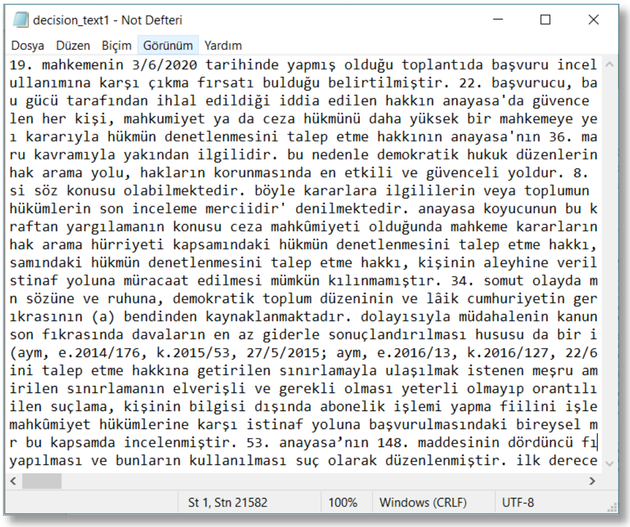


FIGURE 6. The hybrid model structure we created to summarize the decisions of the Constitutional Court of Turkey.

AutoModelForSeq2SeqLM framework, which is influenced by the model's architectural compatibility with sequence-to sequence learning paradigms. The fine-tuning process was governed by Seq2SeqTrainingArguments, which delineates the hyperparameters that are essential for training. The culmination of this training regimen is articulated through the

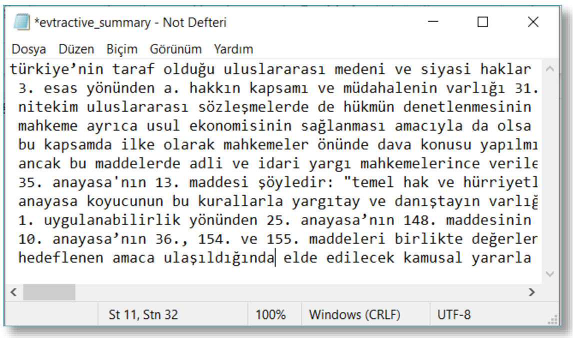


decision_text1 - Not Defteri
Dosya Düzen Biçim Görünüm Yardım

19. mahkemenin 3/6/2020 tarihinde yapmış olduğu toplantıda başvuru incel
ullanımına karşı çıkma fırsatı bulduğu belirtilmiştir. 22. başvurucu, ba
u gücü tarafından ihlal edildiği iddia edilen hakkın anayasa'da güvence
len her kişi, mahkumiyet ya da ceza hükmünü daha yüksek bir mahkemeye ye
ı kararıyla hükmün denetlenmesini talep etme hakkının anayasa'nın 36. ma
ru kavramıyla yakından ilgilidir. bu nedenle demokratik hukuk düzenlerin
hak arama yolu, hakların korunmasında en etkili ve güvenceli yoldur. 8.
si söz konusu olabilmektedir. böyle kararlara ilgililerin veya toplumun
hükümlerin son inceleme merciidir' denilmektedir. anayasa koyucunun bu k
raftan yargılamanın konusu ceza mahkûmiyeti olduğunda mahkeme kararların
hak arama hürriyeti kapsamındaki hükmün denetlenmesini talep etme hakkı,
samındaki hükmün denetlenmesini talep etme hakkı, kişinin aleyhine veril
stinaf yoluna müracaat edilmesi mümkün kılınmamıştır. 34. somut olayda m
n sözüne ve ruhuna, demokratik toplum düzeninin ve lâik cumhuriyetin ger
ıkrasının (a) bendinden kaynaklanmaktadır. dolayısıyla müdahalenin kanun
son fıkrasında davaların en az giderle sonuçlandırılması hususu da bir i
(aym, e.2014/176, k.2015/53, 27/5/2015; aym, e.2016/13, k.2016/127, 22/6
ini talep etme hakkına getirilen sınırlamayla ulaşılmak istenen meşru am
irilen sınırlamanın elverişli ve gerekli olması yeterli olmayıp orantılı
ilen suçlama, kişinin bilgisi dışında abonelik işlemi yapma fiilini işle
mahkûmiyet hükümlerine karşı istinaf yoluna başvurulmasındaki bireysel m
r bu kapsamda incelenmiştir. 53. anayasa'nın 148. maddesinin dördüncü fı
yapılması ve bunların kullanılması suç olarak düzenlenmiştir. ilk derece

St 1, Stn 21582 | 100% | Windows (CRLF) | UTF-8

FIGURE 7. Demonstrative example of the inferential summarization process, showcasing the original decision text. This figure illustrates a document that initially comprises 110 sentences.

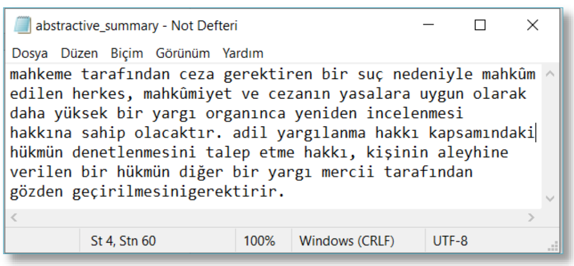


*evtractive_summary - Not Defteri
Dosya Düzen Biçim Görünüm Yardım

türkiye'nin taraf olduğu uluslararası medeni ve siyasi haklar
3. esas yönünden a. hakkın kapsamı ve müdahalenin varlığı 31.
nitekim uluslararası sözleşmelerde de hükmün denetlenmesinin
mahkeme ayrıca usul ekonomisinin sağlanması amacıyla da olsa
bu kapsamda ilke olarak mahkemeler önünde dava konusu yapılmı
ancak bu maddelerde adli ve idari yargı mahkemelerince verile
35. anayasa'nın 13. maddesi şöyledir: "temel hak ve hürriyetl
anayasa koyucunun bu kurallarla yargıtay ve danıştayın varlığ
1. uygulanabilirlik yönünden 25. anayasa'nın 148. maddesinin
10. anayasa'nın 36., 154. ve 155. maddeleri birlikte değerler
hedeflenen amaca ulaşıldığında elde edilecek kamusal yararla

St 11, Stn 32 | 100% | Windows (CRLF) | UTF-8

FIGURE 8. Extractive summary of the original decision text presented in Figure 7. This figure shows how the original text is effectively condensed to a mere 11 sentences post-summarization.



abstractive_summary - Not Defteri
Dosya Düzen Biçim Görünüm Yardım

mahkeme tarafından ceza gerektiren bir suç nedeniyle mahkûm
edilen herkes, mahkûmiyet ve cezanın yasalara uygun olarak
daha yüksek bir yargı organınca yeniden incelenmesi
hakkına sahip olacaktır. adil yargılanma hakkı kapsamındaki
hükmün denetlenmesini talep etme hakkı, kişinin aleyhine
verilen bir hükmün diğer bir yargı mercii tarafından
gözden geçirilmesinigerektirir.

St 4, Stn 60 | 100% | Windows (CRLF) | UTF-8

FIGURE 9. Abstract summary of the original decision text provided in Figure 7. This figure presents a concise overview of the key points from the original decision text.

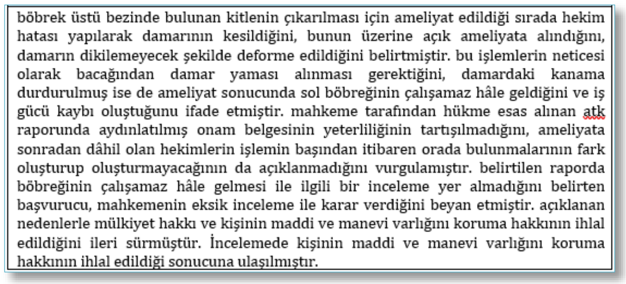
results tabulated in Table 1, demonstrating the performance of the hybrid system in generating summaries.

The efficacy of the hybrid model is corroborated by the performance metrics indicated in Table 1, demonstrating summarization success with Rouge1 and Rouge2 scores of 0.6129 and 0.5884, respectively.

TABLE 1. Hybrid system's performance in generating summaries.

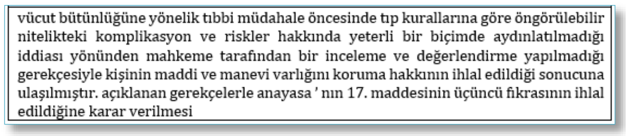
Epoch	Training Loss	Validation Loss	Rouge-1	Rouge-2
1	1.554600	1.269854	0.604700	0.579500
2	1.071000	1.160670	0.607500	0.581400
3	0.910100	1.126786	0.612900	0.588400
4	0.798000	1.118393	0.606400	0.582800

Additionally, Figure 9 presents a manual summary for a "Violation" labeled decision (Decision 2), and Figure 10 illustrates the corresponding summary generated by the hybrid system.



böbrek üstü bezinde bulunan kitlenin çıkarılması için ameliyat edildiği sırada hekim hatası yapılarak damarının kesildiğini, bunun üzerine açık ameliyata alındığını, damarın dikilemeyecek şekilde deforme edildiğini belirtmiştir. bu işlemlerin neticesi olarak bacağından damar yaması alınması gerektiğini, damardaki kanama durdurulmuş ise de ameliyat sonucunda sol böbreğinin çalışamaz hâle geldiğini ve iş gücü kaybı oluştuğunu ifade etmiştir. mahkeme tarafından hükme esas alınan atk raporunda aydınlatılmış onam belgesinin yeterliliğinin tartışılmadığını, ameliyata sonradan dâhil olan hekimlerin işlemin başından itibaren orada bulunmalarının fark oluşturup oluşturmayacağının da açıklanmadığını vurgulamıştır. belirtilen raporda böbreğinin çalışamaz hâle gelmesi ile ilgili bir inceleme yer almadığını belirten başvurucu, mahkemenin eksik inceleme ile karar verdiğini beyan etmiştir. açıklanan nedenlerle mülkiyet hakkı ve kişinin maddi ve manevi varlığını koruma hakkının ihlal edildiğini ileri sürmüştür. İncelemede kişinin maddi ve manevi varlığını koruma hakkının ihlal edildiği sonucuna ulaşılmıştır.

FIGURE 10. A Manual summary for a "Violation" labeled decision 2.



vücut bütünlüğüne yönelik tıbbi müdahale öncesinde tıp kurallarına göre öngörülebilir nitelikteki komplikasyon ve riskler hakkında yeterli bir biçimde aydınlatılmadığı iddiası yönünden mahkeme tarafından bir inceleme ve değerlendirme yapılmadığı gerekçesiyle kişinin maddi ve manevi varlığını koruma hakkının ihlal edildiği sonucuna ulaşılmıştır. açıklanan gerekçelerle anayasa ' nın 17. maddesinin üçüncü fıkrasının ihlal edildiğine karar verilmesi

FIGURE 11. Hybrid system summary for a "Violation" labeled decision 2.

C. DEVELOPMENT OF LEGAL JUDGMENT PREDICTION (LJP) MODELS FOR FULL AND SUMMARIZED TEXTS OF LEGAL DECISIONS

After the summarization process within this research, the implementation of LJP applications, colloquially referred to as legal text prediction, was embarked upon. The primary objective of this phase was to juxtapose the predictive efficacies of models applied to full texts and their summarized counterparts in legal decisions, thereby offering a comparative performance analysis. The predictive system engineered for this purpose comprises five distinct stages. Initially, Turkish legal documents underwent rigorous data preprocessing, which included case normalization, punctuation removal, numerical data exclusion, and elimination of stopwords. The ensuing stage involves constructing a vector space model for each term utilizing the Term Frequency-Inverse Document Frequency (tf-idf) technique, thereby encoding the textual information numerically for subsequent processing. A suite of LJP models comprising XGBoost, MLP, and BERT was developed in the third stage. The Basic Models were subjected to hyperparameter optimization in the fourth stage to increase their prediction performance. The GridSearchCV tool from the scikit-learn library was instrumental in this optimization process, facilitating the systematic exploration and identification of the most efficacious hyperparameter configurations. The optimal hyperparameter values that engendered the highest accuracy in the models are listed in Table 2 for the full-text models and Table 3 for the summary-text models.

TABLE 2. The optimal hyperparameter values for full-text models.

Model	Hyperparameters	Value
XGBoost Model	learning_rate	0.1
	max_depth	3
	n_estimators	500
	subsample	0.6
MLP Model	activation	tanh
	alpha	0.01
	solver	adam
Bert Model	learning_rate	0.001
	optimizer	adam
	Maximum sequence length	512
	Bach size	8

TABLE 3. The optimal hyperparameter values for summary-text models.

Model	Hyperparameters	Value
XGBoost Model	learning_rate	0.01
	max_depth	3
	n_estimators	1000
	subsample	0.8
MLP Model	activation	tanh
	alpha	0.01
	solver	adam
Bert Model	learning_rate	0.01
	optimizer	adam
	Maximum sequence length	380
	Bach size	6

After hyperparameter optimization, the final models were established, and their classification performances were assayed against precision, recall, F1-score, and accuracy metrics. The resulting performance metrics for the models are listed in Table 4 for full-text legal prediction models and in Table 5 for summary-text legal prediction models.

Comparative analysis of the models tailored for predicting "violation" and "non-violation" in full-text legal decisions versus summary-text legal decisions indicates a superior predictive outcome via the XGBoost model. Remarkably, the XGBoost model forecasting full-text legal decisions surpassed the existing literature benchmarks, achieving an accuracy of 93.84%. Moreover, the optimized XGBoost model applied to summary-text legal decisions emerged as the most proficient, with an accuracy of 62.30%.

TABLE 4. Performance metrics for full-text legal prediction.

Model	Class	Precision	Recall	F1-Score	Accuracy
XGBoost Model	non-violation (0)	0.90	0.94	0.92	0.93
	violation (1)	0.95	0.93	0.94	
MLP Model	non-violation (0)	0.80	0.80	0.80	0.80
	violation (1)	0.82	0.80	0.80	
Bert Model	non-violation (0)	0.86	0.84	0.85	0.86
	violation (1)	0.85	0.87	0.86	

TABLE 5. Performance metrics for summary-text legal prediction.

Model	Class	Precision	Recall	F1-Score	Accuracy
XGBoost Model	non-violation (0)	0.55	0.65	0.60	0.62
	violation (1)	0.70	0.60	0.65	
MLP Model	non-violation (0)	0.57	0.60	0.59	0.61
	violation (1)	0.70	0.62	0.64	
Bert Model	non-violation (0)	0.67	0.44	0.53	0.60
	violation (1)	0.57	0.77	0.66	

D. ANNOUNCEMENT OF LJP RESULTS WITH SHAP

Within the scope of the LJP applications conducted in this study, the XGBoost model distinguished itself by delivering the most accurate predictive outcomes. The SHAP method, a prominent technique in explainable artificial intelligence, was employed to elucidate the model's decision-making processes. This approach facilitated a detailed explanation of the factors influencing the model's predictions of summary text legal decisions. A beeswarm plot was utilized to visually represent the variables' importance and marginal contributions to the model's predictions, offering a global perspective on interpretability, as illustrated in Figure 12. This graphical depiction highlights the variable "anayasa" as having the most substantial marginal influence on the predictive outcomes.

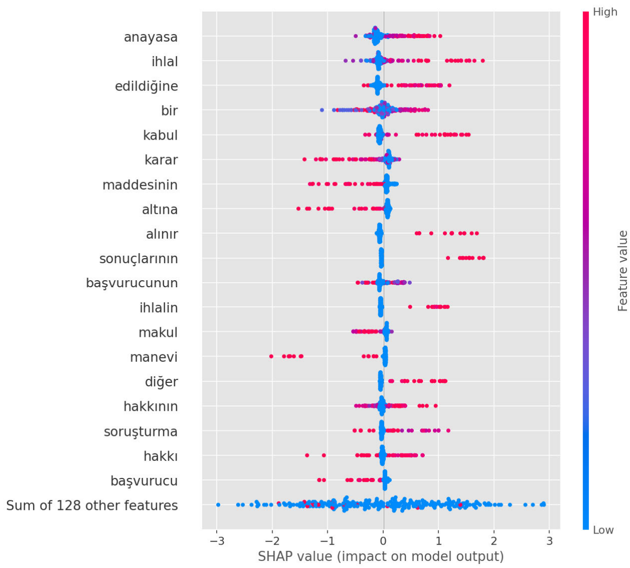


FIGURE 12. Summary text explaining legal decisions with beeswarm plot.

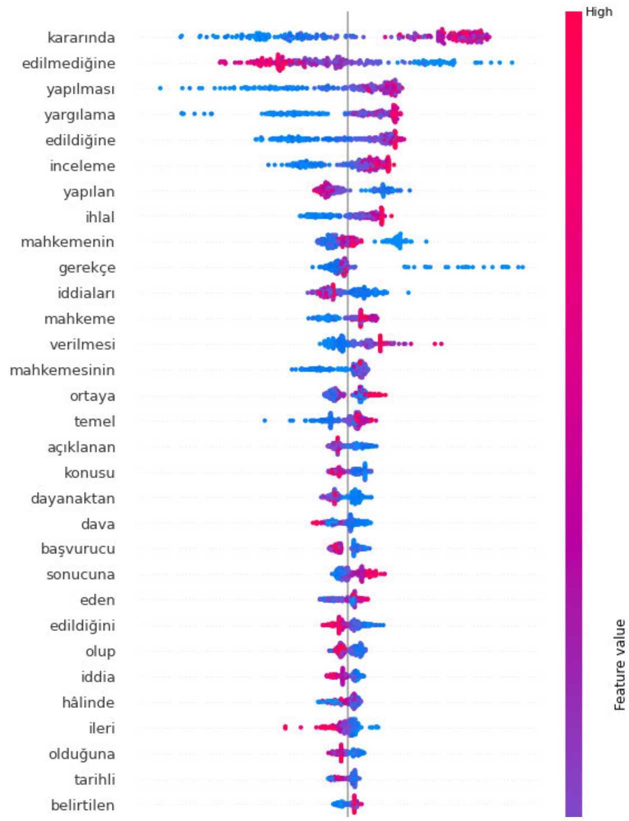


FIGURE 13. Full text explaining legal decisions with beeswarm plot.

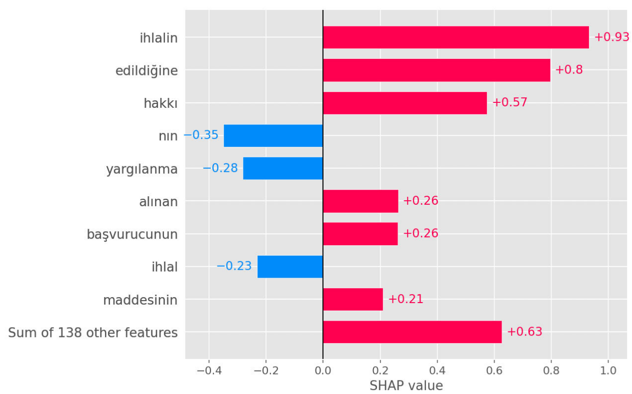


FIGURE 14. Summary text explaining legal decisions with bar plot.

Figure 13 presents a beeswarm plot elucidating the full-text legal decisions, where each plot point signifies an individual case decision, and the X-axis delineates the corresponding SHAP values. The plot underscores the variable ''kararında'' as making the most substantial marginal contribution to the predictions, with an observed trend that increasing values of the ''kararında'' variable are associated with higher SHAP values, thereby indicating a heightened likelihood of a ''Violation'' outcome.

Figure 14 displays a bar chart for the 10th exemplar of summary case decisions, revealing the local interpretability for this specific instance. The chart quantifies the extent to which each variable influences the prediction, with the ''ihlalin'' variable exhibiting the most significant positive impact on the model's output, denoted by a value of ''+0.93''. The predominance of red bars with positive values suggests that, for this particular case, the variables collectively contribute to an increase in the SHAP value, implying a potential ''Violation'' judgment.

Figure 15 delineates a bar chart for the 100th example within the full-text case decisions dataset. This illustration reveals that the variable ''edilmediğine'' exerts the most pronounced negative effect on the f(x) value, marked by a value of ''−1.5''. The prevalence of blue bars bearing negative values indicates that, for this case, the contributing factors collectively diminish the SHAP value, thus inferring a likely ''No Violation'' verdict.

V. DISCUSSION

This study delineated a comprehensive multi-stage approach for analyzing Turkish legal texts, culminating in a series of novel contributions to legal informatics. The inaugural phase involved compiling a bespoke dataset of Turkish legal decisions. This dataset was meticulously constructed through a series of consultations and correspondence with legal experts, thereby ensuring the authenticity and relevance of the compiled texts. Subsequently, the study introduced a hybrid model for legal text summarization, ingeniously blending inferential and abstractive techniques to overcome the inherent limitations of the existing language models. The model's efficacy is substantiated by impressive performance metrics, with Rouge-1, Rouge-2, and Rouge-L scores of 0.6129, 0.5884, and 0.5891, respectively. The outcomes of this novel summarization approach are compared with those from extant literature in Table 6, highlighting the model's superior proficiency.

TABLE 6. Comparing the performances of summarization studies using legal decisions.

	Court	ROUGE-1	ROUGE-2	ROUGE-L
A High-Precision Two-Stage Legal Judgment Summarization [69]	Supreme Court of China	0.4376	0.2711	0.3562
	Federal Court of Australia	0.4315	0.2632	0.3495
Summarization of Indian Legal Judgement Documents via Ensembling of Contextual Embedding based MLP Models [70]	Supreme Court of India	0.64435	0.36228	0.24354
Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation [30]	Supreme Court of UK	0.59	0.41	0.335
	Supreme Court of India	0.575	0.351	0.419
A sentence is known by the company it keeps: Improving Legal Document Summarization Using Deep Clustering [72]	Supreme Court of USA	0.4200	0.2428	03887
	Supreme Court of India	0.5092	0.3599	0.4878
This work	Supreme Court of Turkey	0.6129	0.5884	0.5891

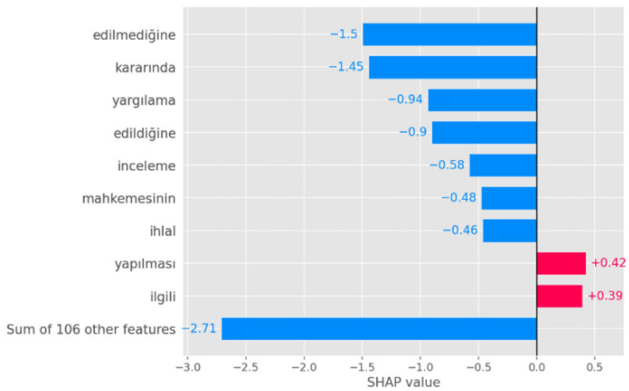


FIGURE 15. Full text explaining legal decisions with bar plot.

TABLE 7. Performance comparison of turkish LJP studies.

	Court	Model	Acc	XAI Method
Using Artificial Intelligence to Predict Decisions of the Turkish Constitutional Court [21]	Supreme Court of Turkey	MLP	90,0%	-
Natural language processing in law: Prediction of outcomes in the higher courts of Turkey [22]	Supreme Court of Turkey	KA	85,1%	-
		RO	87,6%	
		DVM	83,5%	
		DL	91,8%	
This work	Supreme Court of Turkey	XGBoost	93,8%	SHAP
		MLP	80,2%	
		Bert	86,3%	

The third strategic initiative focuses on developing LJP models to predict judicial decisions. A comparative analysis of the success rates and methodologies employed in the LJP models, both within this study and across the broader literature, is detailed in Table 7. The XGBoost model achieved the highest performance, with an accuracy of 93.8%, demonstrating the robustness of ensemble learning techniques in legal judgment prediction. In comparison, BERT and MLP models achieved 86.3% and 80.2% accuracy, respectively, indicating that deep learning models can effectively capture textual features but may require further fine-tuning for optimal legal decision prediction.

In its terminal phase, the study enhanced the interpretability of the most performant XGBoost model by utilizing SHAP visualization techniques, such as beeswarm and bar plots. These techniques provided greater transparency to the model's decision-making processes, making predictive outcomes more accessible and comprehensible.

Despite promising results, this study has several limitations. First, the dataset is limited to Constitutional Court decisions, which may not generalize to other judicial domains such as criminal or administrative law. Future research should consider expanding the dataset to include decisions from various Turkish courts to increase the model's applicability.

Second, while XGBoost demonstrated strong performance, alternative models such as CatBoost and LightGBM should be explored to determine whether further improvements can be achieved. Additionally, transformer-based models, including GPT-4 and T5, could be investigated for more advanced legal text understanding and summarization tasks.

Finally, the computational cost of transformer-based models remains a significant challenge. BERT-based models require substantial resources for training and inference, limiting their deployment in real-time legal applications. Future studies should explore more efficient models, such as DistilBERT or ALBERT, to reduce computational complexity while maintaining high accuracy.

The findings of this study have important implications for legal policy and practice. The proposed AI-based legal judgment prediction and summarization models can serve as valuable decision-support tools for judges, lawyers, and policymakers. By automating the summarization of lengthy legal texts, these models can improve efficiency in legal research and case preparation.

Furthermore, integrating AI-driven decision support systems in judicial processes raises ethical and legal concerns. Ensuring that AI models remain transparent and explainable, as demonstrated through SHAP techniques, is crucial for their acceptance in legal practice. Policymakers should consider regulatory frameworks that guide the ethical use of AI in law, preventing potential biases and ensuring fairness in AI-driven legal applications.

VI. CONCLUSION

This research represents a pioneering endeavor in Turkey to employ AI, complemented by XAI techniques, to predict Constitutional Court decisions and summarize legal texts via a hybrid model. Introducing a new Turkish legal dataset paves the way for future research, where diverse AI algorithms can be deployed to build upon this foundational work, potentially augmenting the corpus of Turkish legal AI applications. The implications of this study are manifold, extending beyond its immediate scholarly contributions to fostering a conducive environment for the growth of Turkish AI research in the legal domain.

It is anticipated that subsequent investigations will leverage the dataset and methodologies presented herein to further the frontier of AI in legal analytics. Future research should focus on:

- Expanding the dataset to include decisions from various Turkish courts to enhance model generalizability.
- Exploring additional AI models, including CatBoost, LightGBM, and GPT-4, to improve predictive performance and legal text understanding.
- Conducting real-world implementation studies to evaluate the practical usability of AI-driven legal decision support systems.
- Enhancing the interpretability of AI models with more advanced XAI techniques beyond SHAP, such as LIME or counterfactual explanations.

By providing a comprehensive framework for summarization and prediction, this study paves the way for future innovations in legal technology and broader application of AI in legal analysis.

REFERENCES

[1] J. Cui, X. Shen, and S. Wen, "A survey on legal judgment prediction: Datasets, metrics, models and challenges," *IEEE Access*, vol. 11, pp. 102050–102071, 2023.

[2] H. Zhong, Z. Guo, C. Tu, C. Xiao, Z. Liu, and M. Sun, "Legal judgment prediction via topological learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Brussels, Belgium, Oct. 2018, pp. 1–11, doi: 10.18653/v1/d18-1390.

[3] S. Long, C. Tu, Z. Liu, and M. Sun, "Automatic judgment prediction via legal reading comprehension," in *Proc. China Nat. Conf. Chin. Comput. Linguistics*, Kunming, China, Oct. 2019, pp. 558–572, doi: 10.1007/978-3-030-32381-3_45.

[4] A. V. Zadgaonkar and A. J. Agrawal, "An overview of information extraction techniques for legal document analysis and processing," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 11, no. 6, p. 5450, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5450-5457.

[5] S. Pandi S, A. M. Farook, W. Kingston, and S. K. Kannaiah, "Enlightening justice: Empowering society through AI-driven legal assistance," in *Proc. 2nd Int. Conf. Adv. Inf. Technol. (ICAIT)*, Jul. 2024, pp. 1–7.

[6] M. Dhore, A. Vimal, A. Agrawal, R. Bajaj, and R. Barde, "Bettercall: AI based legal assistant," in *Proc. 5th Int. Conf. Image Process. Capsule Netw. (ICIPCN)*, Jul. 2024, pp. 248–256.

[7] L. Guo, Q. Liu, K. Shi, Y. Gao, J. Luo, and J. Chen, "A blockchain-driven electronic contract management system for commodity procurement in electronic power industry," *IEEE Access*, vol. 9, pp. 9473–9480, 2021.

[8] M. Muneeb, Z. Raza, I. U. Haq, and O. Shafiq, "SmartCon: A blockchain-based framework for smart contracts and transaction management," *IEEE Access*, vol. 10, pp. 23687–23699, 2022.

[9] N. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos, "Predicting judicial decisions of the European court of human rights: A natural language processing perspective," *PeerJ Comput. Sci.*, vol. 2, p. e93, Oct. 2016, doi: 10.7717/peerj-cs.93.

[10] A. Kaur and B. Bozic, "Convolutional neural network-based automatic prediction of judgments of the European court of human rights," in *Artificial Intelligence and Computer Science*, 2019.

[11] I. Chalkidis, I. Androutsopoulos, and N. Aletras, "Neural legal judgment prediction in english," 2019, *arXiv:1906.02059*.

[12] C. Xiao, H. Zhong, Z. Guo, C. Tu, Z. Liu, M. Sun, Y. Feng, X. Han, Z. Hu, H. Wang, and J. Xu, "CAIL2018: A large-scale legal dataset for judgment prediction," 2018, *arXiv:1807.02478*.

[13] G. Yan, Y. Li, S. Shen, S. Zhang, and J. Liu, "Law article prediction based on deep learning," in *Proc. IEEE 19th Int. Conf. Softw. Qual., Rel. Secur. Companion (QRS-C)*, Sofia, Bulgaria, Jul. 2019, pp. 281–284, doi: 10.1109/QRS-C.2019.00060.

[14] L. Yang, J. Zeng, T. Peng, X. Luo, J. Zhang, and H. Lin, "Leniency to those who confess?: Predicting the legal judgement via multi-modal analysis," in *Proc. Int. Conf. Multimodal Interact.*, Virtual Event, Netherlands, Oct. 2020, pp. 645–649, doi: 10.1145/3382507.3418893.

[15] D. M. Katz, M. J. Bommarito, and J. Blackman, "A general approach for predicting the behavior of the supreme court of the United States," *PLoS One*, vol. 12, no. 4, Apr. 2017, Art. no. e017469, doi: 10.1371/journal.pone.0174698.

[16] B. Strickson and B. De La Iglesia, "Legal judgement prediction for U.K. courts," in *Proc. The 3rd Int. Conf. Inf. Sci. Syst.*, Cambridge, United Kingdom, Mar. 2020, pp. 204–209, doi: 10.1145/3388176.3388183.

[17] G. Semo, D. Bernsohn, B. Hagag, G. Hayat, and J. Niklaus, "ClassActionPrediction: A challenging benchmark for legal judgment prediction of class action cases in the U.S.," 2022, *arXiv:2211.00582*.

[18] M. F. Sert, E. Yıldırım, and İ. Haşlak, "Using artificial intelligence to predict decisions of the Turkish constitutional court," *Social Sci. Comput. Rev.*, vol. 40, no. 6, pp. 1416–1435, Dec. 2022, doi: 10.1177/08944393211010398.

[19] E. Mumcuoğlu, C. E. Öztürk, H. M. Ozaktas, and A. Koç, "Natural language processing in law: Prediction of outcomes in the higher courts of Turkey," *Inf. Process. Manage.*, vol. 58, no. 5, Sep. 2021, Art. no. 102684, doi: 10.1016/j.ipm.2021.102684.

[20] T. Turan, E. Küçüksille, and N. K. Alagöz, "Prediction of Turkish constitutional court decisions with explainable artificial intelligence," *Bilge Int. J. Sci. Technol. Res.*, vol. 7, no. 2, pp. 128–141, Sep. 2023, doi: 10.30516/bilgesci.1317525.

[21] P. Yang, W. Li, and G. Zhao, "Language model-driven topic clustering and summarization for news articles," *IEEE Access*, vol. 7, pp. 185506–185519, 2019.

[22] Y. Du and H. Huo, "News text summarization based on multi-feature and fuzzy logic," *IEEE Access*, vol. 8, pp. 140261–140272, 2020.

[23] I. Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand, and P. K. Soni, "Natural language processing (NLP) based text summarization—A survey," in *Proc. 6th Int. Conf. Inventive Comput. Technol. (ICICT)*, Jan. 2021, pp. 1310–1317.

[24] F. You, S. Zhao, and J. Chen, "A topic information fusion and semantic relevance for text summarization," *IEEE Access*, vol. 8, pp. 178946–178953, 2020.

[25] Z. Altan, "A Turkish automatic text summarization system," in *Proc. IASTED Int. Conf.*, Innsbruck, Austria, Feb. 2004, pp. 1–11.

[26] N. Kemaloglu and E. U. Küçüksille, "Automatic text summarization methods used on Twitter," *Data Sci. Appl.*, vol. 1, no. 1, pp. 9–15, 2021.

[27] S. Polsley, P. Jhunjhunwala, and R. Huang, "Casesummarizer: A system for automated summarization of legal texts," in *Proc. 26th Int. Conf. Comput. Linguistics (COLING)*, Osaka, Japan, Dec. 2016, pp. 258–262.

[28] P. Bhattacharya, S. Poddar, K. Rudra, K. Ghosh, and S. Ghosh, "Incorporating domain knowledge for extractive summarization of legal case documents," in *Proc. 18th Int. Conf. Artif. Intell. Law*, Jun. 2021, pp. 22–31, doi: 10.1145/3462757.3466092.

[29] M. Elaraby and D. Litman, "ArgLegalSumm: Improving abstractive summarization of legal documents with argument mining," 2022, *arXiv:2209.01650*.

[30] A. Shukla, P. Bhattacharya, S. Poddar, R. Mukherjee, K. Ghosh, P. Goyal, and S. Ghosh, "Legal case document summarization: Extractive and abstractive methods and their evaluation," 2022, *arXiv:2210.07544*.

[31] G. Sukanya and J. Priyadarshini, "A meta analysis of attention models on legal judgment prediction system," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, pp. 531–538, 2021, doi: 10.14569/ijacsa.2021.0120266.

[32] S. Li, H. Zhang, L. Ye, X. Guo, and B. Fang, "MANN: A multichannel attentive neural network for legal judgment prediction," *IEEE Access*, vol. 7, pp. 151144–151155, 2019, doi: 10.1109/ACCESS.2019.2945771.

[33] V. Malik, R. Sanjay, S. K. Nigam, K. Ghosh, S. K. Guha, A. Bhattacharya, and A. Modi, "ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation," 2021, *arXiv:2105.13562*.

[34] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, F. Giampaolo, and G. Fortino, "Explainable framework for glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation," *Future Gener. Comput. Syst.*, vol. 129, pp. 152–169, Apr. 2022, doi: 10.1016/j.future.2021.11.018.

[35] B. Uslu and E. U. Küçüksille, "Artificial intelligence library for HTML5 based games: DignityA," *Sakarya Univ. J. Sci.*, vol. 21, no. 1, p. 1, Jan. 2017, doi: 10.16984/saufenbilder.283293.

[36] K. K. Kırboğa, S. Abbasi, and E. U. Küçüksille, "Explainability and white box in drug discovery," *Chem. Biol. Drug Design*, vol. 102, no. 1, pp. 217–233, Jul. 2023, doi: 10.1111/cbdd.14262.

[37] D. Gunning and D. W. Aha, "DARPA's explainable artificial intelligence program," *AI Mag.*, vol. 40, no. 2, pp. 44–58, Jun. 2019, doi: 10.1609/aimag.v40i2.2850.

[38] M. Sharma Timilsina, S. Sen, B. Uprety, V. B. Patel, P. Sharma, and P. N. Sheth, "Prediction of HHV of fuel by machine learning algorithm: Interpretability analysis using Shapley additive explanations (SHAP)," *Fuel*, vol. 357, Feb. 2024, Art. no. 129573.

[39] V. Mizdrakovic, M. Kljajic, M. Zivkovic, N. Bacanin, L. Jovanovic, M. Deveci, and W. Pedrycz, "Forecasting Bitcoin: Decomposition aided long short-term memory based time series modeling and its explanation with Shapley values," *Knowl.-Based Syst.*, vol. 299, Sep. 2024, Art. no. 112026.

[40] A. Aldrees, M. Khan, A. T. B. Taha, and M. Ali, "Evaluation of water quality indexes with novel machine learning and Shapley additive ExPlanation (SHAP) approaches," *J. Water Process Eng.*, vol. 58, Feb. 2024, Art. no. 104789.

[41] M. Todorovic, N. Stanisic, M. Zivkovic, N. Bacanin, V. Simic, and E. B. Tirkolaee, "Improving audit opinion prediction accuracy using metaheuristics-tuned XGBoost algorithm with interpretable results through SHAP value analysis," *Appl. Soft Comput.*, vol. 149, Dec. 2023, Art. no. 110955.

[42] M. Dobrojevic, M. Zivkovic, A. Chhabra, N. S. Sani, N. Bacanin, and M. M. Amin, "Addressing Internet of Things security by enhanced sine cosine metaheuristics tuned hybrid machine learning model and results interpretation based on SHAP approach," *PeerJ Comput. Sci.*, vol. 9, p. e1405, Jun. 2023.

[43] M. A. Kia, "Question-driven text summarization with extractive-abstractive frameworks," Ph.D. dissertation, School of Computer Science, Univ. Essex, Essex, U.K., 2022.

[44] M. Yousefi-Azar and L. Hamey, "Text summarization using unsupervised deep learning," *Expert Syst. Appl.*, vol. 68, pp. 93–105, Feb. 2017, doi: 10.1016/j.eswa.2016.10.017.

[45] R. Nallapati, F. Zhai, and B. Zhou, "Summarunner: A recurrent neural network-based sequence model for extractive summarization of documents," in *Proc. AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, Feb. 2017, pp. 1–14, doi: 10.1609/aaai.v31i1.10958.

[46] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," 2017, *arXiv:1704.04368*.

[47] B. Baykara and T. Güngör, "Turkish abstractive text summarization using pretrained sequence-to-sequence models," *Natural Lang. Eng.*, vol. 29, no. 5, pp. 1275–1304, Sep. 2023, doi: 10.1017/s1351324922000195.

[48] T. Chen, X. Wang, T. Yue, X. Bai, C. X. Le, and W. Wang, "Enhancing abstractive summarization with extracted knowledge graphs and multi-source transformers," *Appl. Sci.*, vol. 13, no. 13, p. 7753, Jun. 2023, doi: 10.3390/app13137753.

[49] T. Ahmed and P. Devanbu, "Few-shot training LLMs for project-specific code-summarization," in *Proc. 37th IEEE/ACM Int. Conf. Automated Softw. Eng.*, Rochester, MI, USA, Oct. 2022, pp. 1–5, doi: 10.1145/3551349.3559555.

[50] F. J. Lozano, D. Alcázar, and A. Díaz, "Web-based application for generation of video-interview summaries using neural networks," in *Proc. 17th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2022, pp. 1–3, doi: 10.23919/CISTI54924.2022.9820457.

[51] A. Gupta, D. Chugh, and R. Katarya, "Automated news summarization using transformers," in *Proc. ICSAC*, Mar. 2021, pp. 249–259, doi: 10.1007/978-981-16-9012-9_21.

[52] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," 2019, *arXiv:1908.08345*.

[53] R. K. Pasumarthi, S. Bruch, X. Wang, C. Li, M. Bendersky, M. Najork, J. Pfeifer, N. Golbandi, R. Anil, and S. Wolf, "TF-ranking: Scalable TensorFlow library for learning-to-rank," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Anchorage, AK, USA, Jul. 2019, pp. 2970–2978, doi: 10.1145/3292500.3330677.

[54] M. Mohd, R. Jan, and M. Shah, "Text document summarization using word embedding," *Expert Syst. Appl.*, vol. 143, Apr. 2020, Art. no. 112958, doi: 10.1016/j.eswa.2019.112958.

[55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, and L. Jones, "Attention is all you need," 2017, *arXiv:1706.03762*.

[56] I. K. Raharjana, D. Siahaan, and C. Fatichah, "User stories and natural language processing: A systematic literature review," *IEEE Access*, vol. 9, pp. 53811–53826, 2021.

[57] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[58] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Proc. Text Summarization Branches Out*, Jul. 2004, pp. 74–81.

[59] S. Song, H. Huang, and T. Ruan, "Abstractive text summarization using LSTM-CNN based deep learning," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 857–875, Jan. 2019, doi: 10.1007/s11042-018-5749-3.

[60] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[61] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 1–11.

[62] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, 2018, doi: 10.1109/ACCESS.2018.2818678.

[63] M. C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Trans. Circuits Syst.*, vol. 8, no. 7, pp. 579–588, 2009.

[64] D. W. Ruck, S. K. Rogers, and M. Kabrisky, "Feature selection using a multilayer perceptron," *J. Neural Netw. Comput.*, vol. 2, no. 2, pp. 40–48, 1990.

[65] R. Cai, B. Qin, Y. Chen, L. Zhang, R. Yang, S. Chen, and W. Wang, "Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM," *IEEE Access*, vol. 8, pp. 171408–171415, 2020.

[66] Ž. Vujovic, "Classification model evaluation metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/ijacsa.2021.0120670.

[67] M. Nazar, M. M. Alam, E. Yafi, and M. M. Su'ud, "A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques," *IEEE Access*, vol. 9, pp. 153316–153348, 2021.

[68] S. Ahmed, M. S. Kaiser, M. Shahadat Hossain, and K. Andersson, "A comparative analysis of LIME and SHAP interpreters with explainable ML-based diabetes predictions," *IEEE Access*, vol. 13, pp. 37370–37388, 2025.

[69] Y. Huang, L. Sun, C. Han, and J. Guo, "A high-precision two-stage legal judgment summarization," *Mathematics*, vol. 11, no. 6, p. 1320, Mar. 2023, doi: 10.3390/math11061320.

[70] D. Jain, M. D. Borah, and A. Biswas, "Summarization of Indian legal judgement documents via ensembling of contextual embedding based MLP models," in *Proc. FIRE Work. Notes*, 2021, pp. 553–561.

[71] D. Jain, M. D. Borah, and A. Biswas, "A sentence is known by the company it keeps: Improving legal document summarization using deep clustering," *Artif. Intell. Law*, vol. 32, no. 1, pp. 165–200, Mar. 2024, doi: 10.1007/s10506-023-09345-y.

TÜLAY TURAN received the bachelor's and first M.S. degrees in electronics and computer education from Suleyman Demirel University, Isparta, Türkiye, in 2008 and 2014, respectively, the second M.S. degree in computer engineering from Pamukkale University, Denizli, Türkiye, in 2015, and the Ph.D. in computer engineering from Suleyman Demirel University, in 2023. Her Ph.D. thesis titled "Legal Text Analysis with Explainable Artificial Intelligence." She is a Faculty Member with the Department of Computer Engineering, Burdur Mehmet Akif Ersoy University, Türkiye. Her research interests include artificial intelligence, machine learning, artificial intelligence ethics, optimization, natural language processing, computer education, and computer science.

ECİR UĞUR KÜÇÜKSİLLE was born in Isparta, Türkiye, in 1976. He received the B.Sc. degree from Gazi University, in 1998, the M.Sc. degree in mechanical education, and the Ph.D. degree in numerical methods from Suleyman Demirel University. From 1998 to 2002, he was with Keçiborlu Vocational School, Suleyman Demirel University, and from 2022 to 2010, he was with the Faculty of Technical Education, Suleyman Demirel University. Since 2010, he has been with the Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Suleyman Demirel University. He is the author of one book and more than 50 articles. He also took part in more than ten projects. He continues to work in the fields of artificial intelligence, machine learning, natural language processing, and cyber security.