International Journal of Religion
2024
Volume: 5| Number 11 | pp. 6791 – 6809
ISSN: 2633-352X (Print) | ISSN: 2633-3538 (Online)
ijor.co.uk
DOI: https://doi.org/10.61707/hqftny98

Enhancing Legal Document Analysis and Judgment Prediction with Machine Learning and Deep Learning Techniques

Avadhut Shelar[1] and Minal Moharir[2]

Abstract

The primary objective of this study is to create a judicial judgement prediction system that achieves a high level of accuracy by employing sophisticated machine learning methods. The goal is to improve the predicting skills of models in legal scenarios by utilising the Texas Wolf Optimisation (TWO) algorithm alongside a Deep Bidirectional Long Short- Term Memory (BiLSTM) network. The approach entails enhancing the BiLSTM model's hyperparameters by the utilisation of the TWO algorithm in order to enhance its capacity to identify intricate patterns within legal documents. The collection consists of previous legal cases from the Supreme Court, which include comprehensive annotations on legal references, arguments, and judgements. Multiple models, such as LR, SVM, CNN, and LSTM, are evaluated for their performance, and the TWO-BiLSTM model demonstrates improved outcomes. Models are evaluated using performance criteria such as accuracy, F-score, precision, and recall. The findings demonstrate that the TWO-BiLSTM model has superior performance compared to current models, with a 97% accuracy and a 97.29% F-score in scenarios with a True Positive (TP) rate of 90. Furthermore, it consistently demonstrates robust performance in K-fold cross-validation, with an impressive accuracy rate of 96%. The study showcases the efficacy of the suggested TWO-BiLSTM model as a robust tool for forecasting judicial outcomes, presenting significant enhancements compared to conventional methods.

Keywords: *Legal Judgment Prediction, BiLSTM, Texas Wolf Optimization (TWO) Machine Learning and Hyperparameter Optimization*

INTRODUCTION

The swift advancement of technology has greatly influenced different industries, including the legal field. The incorporation of machine learning (ML) and deep learning (DL) methods into legal informatics signifies a notable progress, with the potential to transform the analysis of legal documents and the prediction of court rulings. The objective of this study, named "Enhancing Legal Document Analysis and Judgement Prediction with Machine Learning and Deep Learning Techniques," is to investigate the efficacy, constraints, and wider ramifications of these technologies in the legal domain. The utilisation of machine learning (ML) and deep learning (DL) in the field of legal informatics is motivated by the necessity to effectively handle and analyse extensive quantities of legal data. Conventional approaches of analysing legal documents are frequently lengthy and susceptible to mistakes made by humans. By harnessing the power of machine learning (ML) and deep learning (DL), it is feasible to automate a significant number of these processes, resulting in outcomes that are more precise and consistent[1]–[9]. The research has the potential to revolutionise legal practices by improving efficiency, accuracy, and decision-making processes. This study explores a range of machine learning models, encompassing both traditional algorithms and sophisticated deep learning architectures, to assess their effectiveness in analysing legal documents and predicting judgements[10]–[17]. More precisely, it evaluates various models including Logistic Regression, Support Vector Machines (SVM), Random Forest, and advanced deep learning architectures such as Long Short-Term Memory (LSTM) networks and Bidirectional LSTM

(BiLSTM) networks. Every model undergoes thorough evaluation to assess its precision, recall, F1-score, and other pertinent metrics in order to ascertain its appropriateness for particular legal tasks. A primary area of concentration is the anticipation of court rulings, which is an intricate undertaking that necessitates comprehension of the intricacies of legal terminology and context. The study investigates the manner in which various machine learning models manage the complexities of legal texts and their ability to accurately forecast outcomes using historical data. The results emphasise that whereas typical machine learning models perform

[1] Research Scholar, Department of CSE, RVCE, Bangalore and Visvesvaraya Technological University, Belagavi –590018, India

[2] Professor and Research Supervisor, Department of CSE, RVCE, Bangalore and Visvesvaraya Technological University, Belagavi - 590018, India

satisfactorily, deep learning architectures, specifically LSTM and BiLSTM networks, provide greater predictive accuracy. The models' capacity to comprehend sequential relationships and contextual information in legal texts renders them exceptionally efficient for predicting judgements. Nevertheless, the efficacy of these models varies in different contexts. The efficacy of machine learning (ML) and deep learning (DL) models can exhibit substantial variation depending on the characteristics of the legal documents and the particular setting of the cases. The variability highlights the significance of choosing models that are in line with the specific job and the peculiarities of the dataset[18]–[23]. The report offers a comprehensive analysis of these parameters, providing valuable insights into the optimisation of alternative models for diverse legal purposes. An additional pioneering element of this study involves the investigation of the Texas Wolf Optimizer (TWO), which is an optimisation algorithm that draws inspiration from the social structure and hunting tactics employed by Texas wolves. The use of TWO is applied to optimise the performance of deep learning models, namely in the field of legal document processing and prediction of judgements. The approach enhances the convergence speed and accuracy of deep learning models by optimising hyperparameters. The study showcases the potential of mixing evolutionary algorithms with deep learning (DL) for difficult legal problems by integrating TWO with BiLSTM networks, resulting in notable enhancements in prediction performance. This research not only focuses on predicting judgements but also investigates the creation of the Smart Law Annotator, a tool specifically developed to automate the process of annotating legal texts. Annotation is an essential procedure in the examination of legal documents, which entails the identification and categorisation of crucial information within texts. The Smart Law Annotator employs natural language processing (NLP) and deep learning (DL) techniques to optimise this procedure, resulting in a substantial decrease in the time and effort expended by legal professionals. The architecture of the programme combines sophisticated natural language processing (NLP) techniques with deep learning (DL) models to achieve exceptional precision and efficiency in annotating documents. The practical consequences of this research have a wide and significant impact[24]–[30]. ML and DL technologies can boost the operational efficiency of legal practitioners by automating document analysis and improving judgement prediction. This allows legal professionals to concentrate on more intricate and subtle areas of legal practice. This transition has the capacity to decrease expenses, enhance the efficiency of legal procedures, and elevate the general calibre of legal services. Moreover, the study emphasises the moral and logistical factors linked to using machine learning (ML) and deep learning (DL) in the legal field. The examination critically evaluates issues such as data privacy, model transparency, and the possibility of bias. The study promotes the establishment of strong ethical principles and regulatory frameworks to control the utilisation of these technologies in legal practice. To summarise, this research highlights the significant capacity of machine learning (ML) and deep learning (DL) in the field of legal informatics. By offering a thorough examination of several models and their uses in the study of legal documents and the prediction of judgements, it establishes the foundation for future progress in this area. The knowledge acquired from this research is anticipated to guide the creation of more advanced instruments and approaches, ultimately enhancing and modernising legal procedures. With the increasing adoption of technology in the legal field, the incorporation of machine learning (ML) and deep learning (DL), as well as optimisation methods such as the Texas Wolf Optimizer, will have a significant impact on the future of legal services. These advancements will enhance the efficiency, accuracy, and accessibility of legal services[31]–[38].

LITERATURE REVIEW

Valvoda 2023 et.al the law in one of the following two ways. It either expands its scope, in which case it sets positive precedent, or it narrows it, in which case it sets negative precedent. Legal outcome prediction, the prediction of positive outcome, is an increasingly popular task in AI. In con-trast, we turn our focus to negative outcomes here, and introduce a new task of negative outcome prediction. We discover an asym-metry in existing models' ability to predict positive and negative outcomes. Where the state-of-the-art outcome prediction model we used predicts positive outcomes at 75.06 F1, it predicts negative outcomes at only 10.09 F1, worse than a random baseline. To address this performance gap, we develop two new models inspired by the dynamics of a court process. Our first model significantly improves positive outcome prediction score to 77.15 F1 and our second model more than doubles the negative outcome prediction performance to 24.01 F1. Despite this improvement, shifting focus to negative outcomes reveals that there is still much room for improvement for outcome prediction models [39].

Cui 2023 et.al Legal judgement prediction (LJP) automatically predicts judgement findings from fact descriptions by applying Natural Language Processing (NLP) techniques. The increasing curiosity about using natural language processing methods for LJP is the driving force behind the current effort. The availability of large-scale public datasets and recent developments in natural language processing have led to encouraging findings on several benchmark datasets, despite the present performance gap between humans and robots. The following are some of the contributions made by this study to the current state of LJP tasks, datasets, models, and evaluations: 4) state-of-the-art results for 11 representative datasets from different court cases and an in-depth discussion of the open challenges in this area. 5) an analysis of 43 LJP datasets constructed in 9 different languages, along with a LJP classification method based on three attributes. 6) a summary of 16 evaluation metrics categorised into 4 types to evaluate the performance of LJP models for different outputs. 7) a review of 8 legal- domain pretrained models in 4 languages, highlighting four major research directions for LJP.

8) a review of 8 legal-domain pretrained models in 4 languages, highlighting four major research directions for LJP. 6) state-of-the-art results for the datasets and an in-depth discussion of the open challenges in this area. Researchers in natural language processing (NLP) and the legal field can use this study's extensive overview of LJP's recent developments to better understand the field, and they can work together to improve LJP models' performance [5].

Fei 2023 et.al A number of areas have shown that large language models (LLMs) are very capable. Their level of legal expertise and reliability in handling legal-related duties are questions that arise when applied to the highly specialised, safe-critical legal area. We suggest a thorough evaluation standard, LawBench, to fill this void. Careful design went into making sure that LawBench accurately assessed LLMs' legal competence on three different cognitive levels: (3) Legal knowledge applying: whether LLMs can appropriately apply their legal knowledge and make the necessary reasoning steps to solve realistic legal tasks; (4) legal knowledge understanding: whether LLMs can comprehend entities, events, and relationships within legal text; and (5) legal knowledge memorisation: whether LLMs can memorise needed legal concepts, articles, and facts. There are a total of twenty tasks in LawBench, spanning five different types of tasks: generation, regression, single-label classification (SLC), and multi- label classification (MLC). We rigorously assess 51 LLMs on LawBench, comprising 20 bilingual LLMs, 22 LLMs with an emphasis on China, and 9 LLMs with a focus on law. The results demonstrate that, when compared to other LLMs in the legal arena, GPT-4 continues to perform exceptionally well. We still have a ways to go before we get LLMs that are both useable and dependable for legal tasks, even though refining them on legal specific material does offer some advances. Check out https://github.com/open-compass/LawBench/ for all the data, model predictions, and evaluation code. With any luck, this benchmark will help shed light on the LLMs' domain-specific skills and hasten their development for use in the legal field [4].

Dhanani 2023 et.al In order to find comparable decisions and prepare beneficial and strategic arguments for the Court, legal experts are vehemently in favour of an automated and user- friendly legal document recommendation system (LDRS). Doc2Vec does a fantastic job at learning vector space, which contains embeddings with rich semantic information, from the judgement corpus text. Applying previous domain-specific information during Doc2Vec learning has the ability to improve the embedding representation. This study therefore suggests a legal domain-specific pre-learned word embedding LDRS (P-LDRS) that learns the Doc2Vec embedding with the semantic knowledge of the legal domain. Doc2Vec runs into scalability problems when trying to learn judgement embedding from large legal documents that already exist. With the help of frameworks like MapReduce and Spark, the suggested P-LDRS may learn the judgement embedding distributedly over a cluster of computing nodes, which solves the scalability problem. A distributed and a non-distributed version of the suggested P-LDRS are both tested empirically to verify its efficacy and scalability. With an Accuracy of 0.88, F1- Score of 0.82, and MCC Score of 0.73, the experimental results demonstrate that the suggested non-distributed P-LDRS outperforms the classic Doc2Vec based LDRS by a substantial margin. As the number of nodes increases, they show that the proposed distributed P-LDRS improves time efficiency and achieves consistent Accuracy of ≈0.88, F1-Score of ≈0.83, and MCC Score of ≈0.72 [25].

Trautmann 2023 et.al Legal Prompt Engineering, also known as Legal Prompting, is a method for training and assisting large language models (LLMs) to carry out NLLP skills. For the purpose of the Legal Judgement Prediction (LJP) job, we intend to apply LPE with LLMs throughout lengthy legal documents. We study how well zero-shot LPE works with provided facts in case-texts from the Swiss Federal Supreme Court (in Italian, German, and French) and the European Court of Human Rights (in English). While zero-shot LPE outperforms the baselines, it is not up to par with the most cutting-edge supervised methods available today, according to our findings. Our findings demonstrate that general-purpose LLMs can be applied to the legal domain without the need for explicit domain-specific data. Additionally, the LLMs were applied directly without additional training or fine-tuning, resulting in significant savings in computational costs [18]

Author/year	Method	Research gap	Controversies
Lyu/2022 [9]	"New RL-based framework CEEN for improved LJP by the extraction of discriminatory criminal aspects."	"Regarding LJP papers that portray facts indistinguishably and those that present deceptive legal arguments."	"Legal ambiguities and incorrect conclusions in LJP (automated)"
Kalamkar/2022 [40]	"New database of entities with legal status and initial model for dataextraction."	"Present datasets fail to identify legal named entities with sufficient granularity."	"Difficulties in specifying and uniformly naming certain legal entitiesat the micro level."
Feng/2022 [41]	"Statement of the past, present, and future of LJP milestones in manylanguages and jurisdictions."	"Several jurisdictionsand languages do nothave uniform LJP models."	"Controversies regarding the reliability and generalizability of LJP in different legal systems."
Hwang/2022 [22]	"Korean legal datasets and a linguistic model applied to varioustasks for benchmarking."	"Disruptive legal natural language processing challenges: a lack oflarge-scale, non- English legal datasets."	"Worries regarding the coverage of non-English languages inthe dataset, potentialbias, and limitations in the dataset's breadth."
Malik/2021 [30]	"ILDC presents Explainability Models for CourtJudgement Prediction and Explanation."	"Automated Court Judgement Prediction systemshave limited explainability and accuracy."	"There are real- world implications to the fact that algorithmic and expert explanationsdiffer."

METHODOLOGY

The methodology commences with the utilisation of the Texas Wolf Optimisation (TWO) algorithm, which draws inspiration from the hierarchical and cooperative dynamics observed in wolf packs. This algorithm is employed to address intricate optimisation problems. TWO begins by constructing a hierarchical pack structure, in which alpha, beta, and delta wolves lead the quest for the best answers. The method achieves a balance between exploration and exploitation by emulating the hunting strategies of wolves, resulting in enhanced solution quality and convergence. BiLSTM networks improve the processing of sequential data by including information from both past and future contexts, hence enhancing performance in applications like as language modelling and time-series prediction. The Smart Law Annotator utilises powerful Natural Language Processing (NLP) and deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to extract crucial information and guarantee adherence to regulations, thereby streamlining the study of legal documents. The procedure entails preprocessing unprocessed texts, annotating them with pertinent information, training models using labelled data, and assessing performance using test data. Active learning enhances the model by repeatedly integrating additional labelled data. Utilising JSON schemas for legal material and automated reasoning guarantees accurate and current legal analysis, hence improving operational efficiency and ensuring compliance in legislative systems. This comprehensive methodology integrates TWO's optimisation, BiLSTM's data processing, and advanced legal text analysis to successfully tackle real-world difficulties.

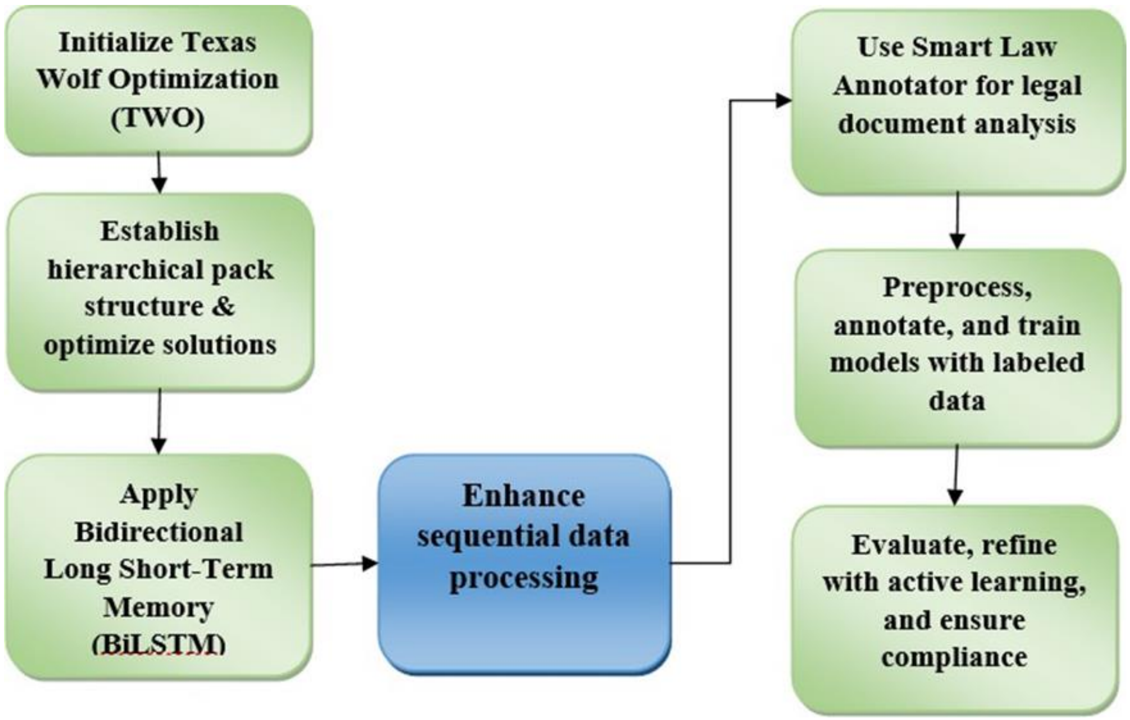


Figure 1 Proposed Flowchart

Texas Wolf Optimization (TWO)

Texas Wolf Optimisation (TWO) is an advanced optimisation algorithm that draws inspiration from the social dynamics and hunting methods of Texas wolves. This method is a component of nature-inspired optimisation strategies that utilise animal behaviour and principles of biological evolution to efficiently address complicated optimisation issues. TWO explores the solution space by simulating the hierarchical and cooperative organisation of a wolf pack, which includes alpha leaders, beta subordinates, and omega followers. The algorithm tackles obstacles such as local optima, large dimensionality, and non-linearity by guiding the search process towards solutions that are close to optimal. TWO offers a strong framework for addressing real-world optimisation problems by imitating the cooperative hunting behaviour of wolves. It outperforms existing algorithms in terms of both convergence time and solution quality.

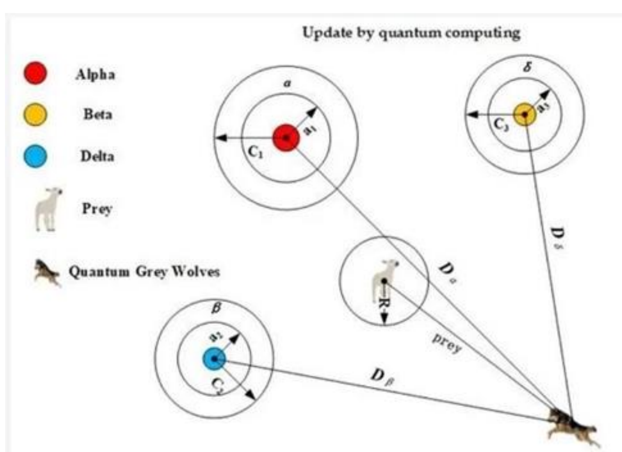


Figure 2 Texas Wolf Optimizer

Core Principles and Mechanisms of TWO

TWO is guided by a set of fundamental rules that are based on the behaviour of wolves. The initial step involves the establishment of a pack structure and a hierarchical system of leadership, wherein a set of viable solutions is created and assessed based on a fitness function. The most successful solutions, referred to as alpha, beta, and delta wolves, create a hierarchical structure that directs the search process. The second premise is based on the concepts of exploration and exploitation, which are inspired by the hunting strategies employed by wolves. The programme achieves a balance between enhancing current solutions (exploitation) and exploring new areas (exploration) by guiding the pack towards promising places identified by the alpha, beta, and delta wolves. The third concept involves encircling prey, which entails updating the positions of prospective solutions using mathematical models that mimic the encircling behaviour of wolves. This approach helps sustain population variety while also ensuring convergence. The ultimate principle involves pursuing and engaging with prey, accelerating the exploration of the most favourable alternatives, similar to the final attack of a wolf pack on its victim.

Applications and Performance of TWO

The TWO algorithm has demonstrated successful use in a diverse array of optimisation issues spanning multiple fields. TWO, in the field of engineering design, enhances the performance and efficiency of mechanical components, structural designs, and control systems. Operational research deals with the optimisation of scheduling, routeing, and resource allocation problems, offering solutions that are both highly effective and efficient. Machine learning use TWO for the purposes of parameter adjustment, feature selection, and model optimisation, hence improving algorithm performance. TWO's effectiveness is frequently compared to that of other optimisation methods, such as Genetic methods (GA), Particle Swarm Optimisation (PSO), and Grey Wolf Optimizer (GWO). TWO often exhibits higher performance, especially in intricate and multi-modal environments where alternative algorithms may encounter difficulties. The durability, simplicity, and adaptability of TWO make it a great tool for researchers and practitioners in diverse domains.

Texas Wolf Optimization (TWO)

```
Texas Wolf Optimization (TWO) Initialize population of wolves
(solutions)
Define fitness function to evaluate solutionsSet alpha, beta, and
delta wolves

While stopping criteria not met do: For each wolf in the
    population do:
        Evaluate fitness of the wolf
        Update alpha, beta, and delta wolves based on fitnessCalculate new positions of
        wolves:
            For each wolf do:
                Update position using encircling prey strategyBalance exploration and
                exploitation
    End for
End while

Output best solution found
```

BiLSTM (Bidirectional Long Short-Term Memory) Overview

BiLSTM networks are a sophisticated type of recurrent neural networks (RNNs) that are specifically designed to improve the processing of sequential input. BiLSTMs, in contrast to standard RNNs, employ two LSTM networks that operate in opposite directions: one processes the input sequence in a forward manner, while the other processes it in a backwards manner. This bidirectional technique allows the model to gather contextual information from both previous and next states, overcoming the constraints of unidirectional LSTMs. BiLSTMs enhance performance in applications such as language modelling, speech recognition, and time-series prediction by combining outputs from both directions at each time step. The forward LSTM model collects dependencies from the preceding context, while the backwards LSTM model captures dependencies from the subsequent context. The concatenation layer merges the hidden states from both LSTMs to generate a comprehensive representation at each time step, hence greatly improving the ability to process sequential data.

Pseudo code of Bidirectional Long Short-Term Memory (BiLSTM)
Bidirectional Long Short-Term Memory (BiLSTM) Initialize BiLSTM network with forward and backward LSTMsFor each input sequence do: Forward LSTM processes sequence from start to end Backward LSTM processes sequence from end to startConcatenate outputs of forward and backward LSTMsApply dropout to prevent overfitting Pass through dense layer with softmax activation for classificationEnd for Train network with labeled data Evaluate performance with test data

Legal Document Analysis and Deep Learning Applications

Legal document analysis is a systematic process that carefully examines and interprets legal texts in order to extract important information, guarantee adherence to regulations, and provide assistance for well-informed decision-making. This procedure utilises sophisticated methodologies such as Natural Language Processing (NLP) and machine learning algorithms to effectively manage the intricacy and extent of legal documents, including contracts, regulations, and court judgements. Natural Language Processing (NLP) enables the identification and retrieval of important terms, clauses, and entities, which are crucial for activities such as contract analysis and due diligence. Machine learning algorithms, encompassing both supervised and unsupervised techniques, are capable of categorising texts, recognising patterns, and generating predictions by leveraging past data. Deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), improve this procedure by identifying patterns and capturing contextual connections. Utilising modern deep learning architectures such as BiLSTM enhances the comprehension and analysis of legal texts, leading to more effective and precise administration of legal documents. These technologies optimise the process of reviewing documents, decrease the time required for analysis, and enhance accuracy, thus becoming essential in the legal field.

Pseudo code Legal Document Analysis with Deep Learning
// Legal Document Analysis with Deep Learning Preprocess raw legal texts (cleaning, segmentation)Generate unlabeled judgments from raw data Annotate judgments with relevant information Create labeled dataset for training Initialize deep learning model (e.g., CNN, RNN, BiLSTM)Train model with labeled data Evaluate model performance with test dataApply active learning to refine model: Incorporate new labeled dataUpdate and retrain model Use JSON schemas to manage and interpret legal data Ensure compliance and accuracy with automated reasoning Output analysis and recommendations

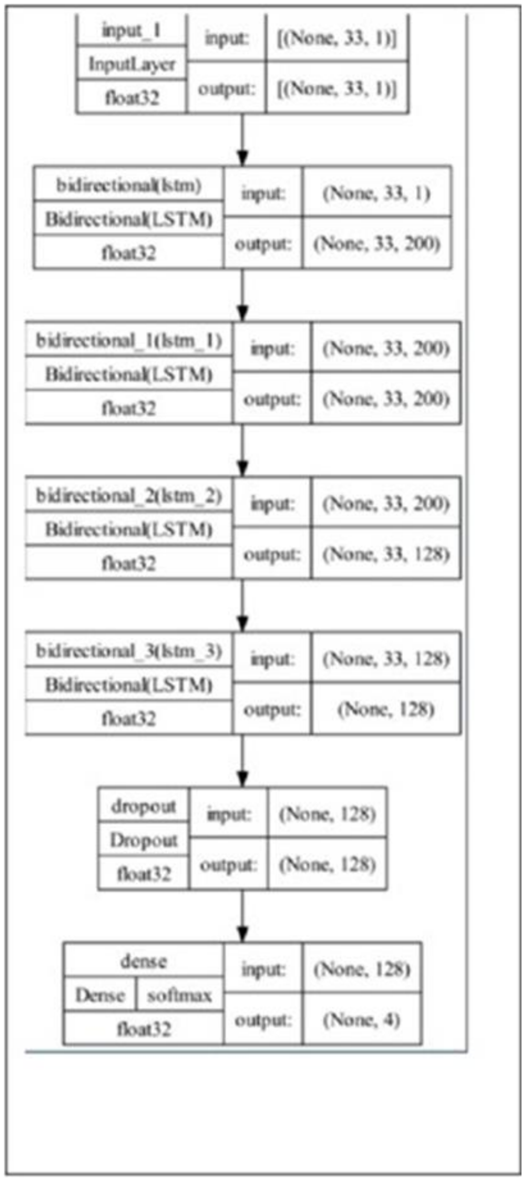


Figure 3 BiLSTM model architecture

The figure depicts the architectural design of a Bidirectional Long Short-Term Memory (BiLSTM) model, which is an advanced technology utilised in the field of machine learning. This architectural design starts with input layers that transmit sequential data into bidirectional LSTM layers. These layers possess a distinctive characteristic in that they are capable of processing data in two directions: forward and backwards. This allows them to successfully capture patterns from both previous and forthcoming points in sequences. This bidirectional method enables the model to gain a more full understanding of the context, hence improving its capacity to analyse and make predictions using sequential input. In order to mitigate the problem of overfitting, dropout layers are included in the model architecture. These layers randomly disable a portion of the neurones throughout the training process, hence promoting better generalisation of the model to unseen data. The last phase of the architecture comprises a dense layer that is fitted with softmax activation. The inclusion of this layer is essential for classification jobs as it transforms the processed data into probability distributions across several classes. BiLSTMs are highly effective in natural language processing (NLP) and time series analysis due to their capacity to capture bidirectional context, rendering them potent for tasks such as language modelling, speech recognition, and sentiment analysis. By including input from both forward and backwards directions, this architecture greatly improves the ability to model sequences, resulting in more precise and dependable predictions and classifications.

Smart Law Annotator and Deep Learning for Legal Document Analysis

The Smart Law Annotator is an advanced tool that utilises machine learning to improve the speed and precision of legal document processing. The system utilises sophisticated NLP methods to extract important information, identify crucial terms and clauses, and guarantee adherence to regulatory standards. The Smart Law Annotator utilises deep learning models such as CNNs and RNNs to automate the analysis and interpretation of legal texts. This connection streamlines duties such as contract analysis, due diligence, and regulatory compliance. The application utilises deep learning architectures to offer legal practitioners valuable insights and recommendations, leading to substantial enhancements in decision- making processes and a decrease in the likelihood of human error.

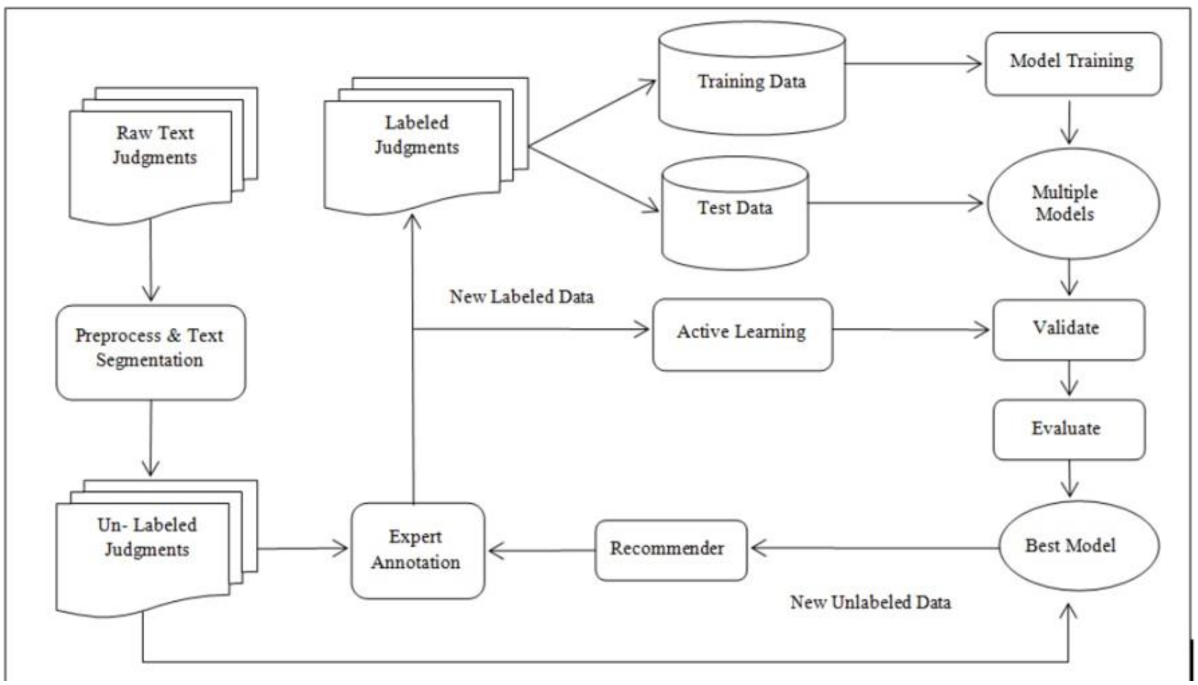


Figure 4 an overview of the process of machine learning and annotation

The flowchart illustrates the complex procedure of machine learning and annotation, which is essential for the development of precise and dependable models. The process commences with "Raw Text Judgements," which are initially subjected to "Preprocess & Segmentation." This stage entails the process of cleansing and disassembling the text into smaller, more digestible chunks, in order to facilitate subsequent analysis. Next, we have "Un-Labeled Judgements," which refers to raw data that does not have any annotations. Afterwards, skilled annotators intervene to generate "Labelled Judgements," in which every data point is marked with pertinent information, so converting it into a valuable asset for training purposes. The labelled data points are crucial for "Model Training," as they are combined with additional "Training Data" to instruct the machine learning model in identifying patterns and making predictions. In order to ascertain the model's efficacy, its performance is meticulously assessed using "Test Data," which the model has not been exposed to previously. This aids in evaluating its precision and ability to apply to a wide range of situations. The model's performance is regularly improved through an iterative cycle called "Active Learning," in which new labelled data is used. The recommender system plays a crucial role in this process by providing new unlabeled data that experts can annotate. The collaborative interaction between expert annotation and machine learning is a dynamic process that continuously improves the model's accuracy and dependability.

JSON Schemas for Legal Content combine rule-based systems with machine learning approaches, leading to a significant transformation in the process. The Legal Rule ML system improves the effectiveness of legal analysis, decision-making, and automated compliance checks. Legal Rule ML enhances the uniformity and precision of legal document depiction, aiding legal practitioners in understanding, implementing, and enforcing laws. The technique adjusts to changing legislation and legal norms, guaranteeing that representations stay up-to- date and accurate. Moreover, the use of automated reasoning enhances trust in the generated analyses due to its transparency. Legal Rule ML enhances operational efficiency, accuracy, and compliance with legal standards, hence revolutionising conventional legal procedures.

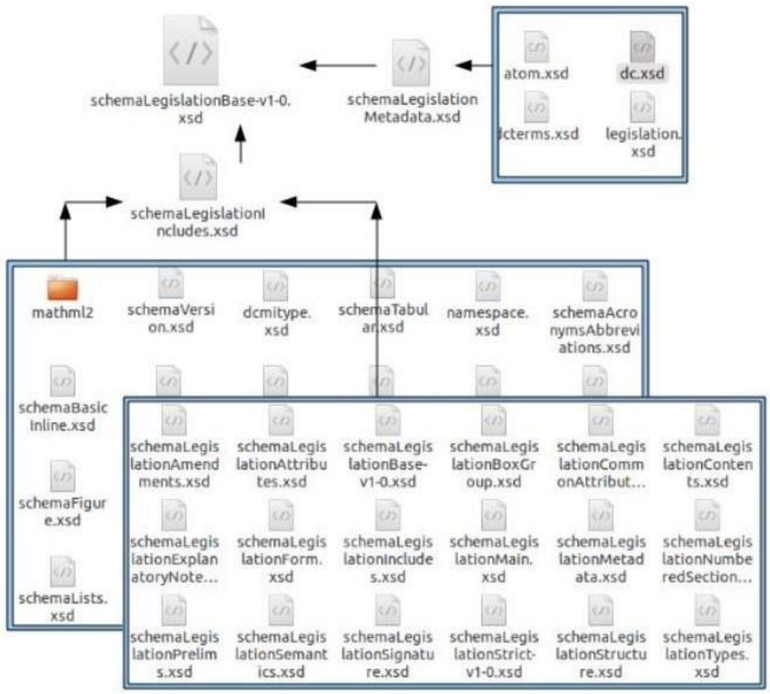


Figure 5 schemaLegislationBase-v1.0.xsd

The figure illustrates an intricate arrangement of XML schema files for a legislative system, namely version 1.0.0. The file "schemaLegislationBase-v1.0.0.xsd" is the central component of the complex web of schema architecture. The primary schema effortlessly links to other essential schemas, such as "atom.dc.xsd," "bcterms.xsd," and "legisld.xsd." Adjacent to these core schemas are supplementary schema files, each intricately designed to serve a distinct purpose inside the legislative information system. The arrows connecting these schema files represent relationships or dependencies, illustrating how modifications in one schema could affect others within the network. The integrated structure guarantees efficient and uniform management of legislative data throughout the system. The different schemas collaborate to facilitate the efficient management of legislative papers, metadata, and words. The graphic depicts a complex and orderly architecture for handling legislative data, showcasing the system's ability to uphold precision, uniformity, and thorough data management through its intricate and interrelated schema files. This framework is essential for legislative data systems, serving as a foundation for encoding, analysing, and applying legislative texts accurately and dependably.

RESULT & DISCUSSION

Performance Evaluation

Table 1 Analysed LR, SVM, CNN, and LSTM utilising Precision (P), Recall (R), and F1 Score (F1).

	Allowed			Rejected		
Method	P	R	F1	P	R	F1
TFIDF-LR	0.00	0.00	0.00	0.78	1.00	0.88
TFIDF-SVM	0.87	0.41	0.56	0.86	0.98	0.92
CNN-W2V-EMB	1.00	0.11	0.21	0.80	1.00	0.89
LSTM-W2V-EMB	0.22	0.12	0.15	0.78	0.88	0.83
LSTM-Kera-EMB	0.87	0.82	0.86	0.95	0.97	0.96

Table 1 displays the evaluation of LR, SVM, CNN, and LSTM models based on Precision (P), Recall (R), and F1 Score (F1). The table provides a comparative analysis of four machine learning models (LR, SVM, CNN, and LSTM) with regards to their performance on a classification task. The table presents the precision, recall, and F1 score for each model. Precision is a metric that calculates the percentage of relevant objects that a model correctly classified, while recall is a metric that calculates the percentage of relevant things that the model accurately identified. The F1 score is a metric that combines precision and recall in a balanced manner, derived using a weighted harmonic mean. Higher numerical numbers are preferable. As an example, the CNN model demonstrates the highest precision, achieving a value of 1.00. However, its recall is the lowest, measuring at 0.11.The text is enclosed in a pair of asterisks. This table presents a succinct comparison of the performance of four machine learning models in a categorisation test.

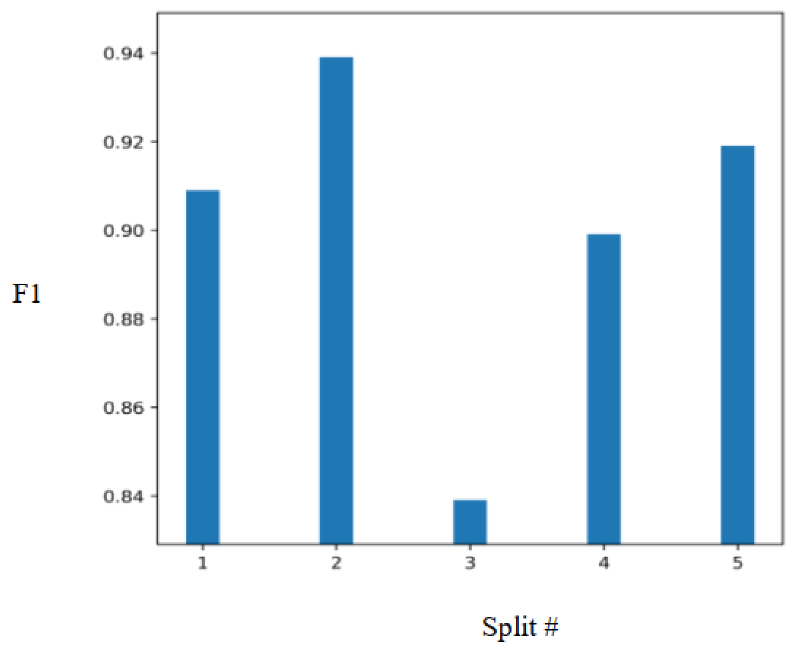


Figure 6 Very good f1-scores in different splits.

Figure 6 exhibits outstanding F1-scores across multiple divisions. The F1 score, a statistic used to assess a model's performance in a classification test, is represented on the vertical axis. Higher scores indicate higher performance. The x-axis represents several subdivisions of the data. All data divisions have a good F1-score (above 0.90), as indicated by the bars. This suggests that the model has exceptional proficiency in this specific task. The graph clearly illustrates the exceptional performance of a machine learning model on a classification task, as assessed by the F1-score metric, with a concise representation of only 49 words.

Table 2 Inter annotator agreement

User	a1	a2	a3
a1	**1.0000**	**0.9210**	**0.9112**
a2	**0.9210**	**1.0000**	**0.9291**
a3	**0.9112**	**0.9291**	**1.0000**

Table 2 displays the level of agreement between annotators. The data presents the collective agreement of three annotators (al, a2, and a3) for a certain task. Each entry in the table represents the ratio of annotations when two annotators agreed on a decision. The value 0.9210 in the cell at row al and column a2 indicates that annotators al and a2 agreed on 92.10% of the annotations. An optimal level of agreement would be denoted by a numerical value of 1.0000. Based on the information provided in the table, it can be observed that the three annotators demonstrate a notable level of agreement.

Table 3 Statistics of the labeled dataset.

Element Type		Tokens	Intstances
Legal References		2221	23612
Argument by	Applicant RespondentJudge	2958 3187 4768	30480 39873 84322
Argument Sentence Type	Premise Conclusion	734 463	8425 1243
Order Type	Accepted Rejected	1113 109	19811 554

The table presents statistical data regarding a labelled dataset. The paper presents a thorough breakdown of the number of tokens and instances for different elements, including legal references, arguments put up by the applicant or respondent, and judgements. For example, the dataset contains a total of 2221 tokens that reflect legal references. Furthermore, these legal references are included a grand total of 23612 times across the dataset.

The TWO-BiLSTM model is utilised to develop a legal judgement prediction system, and its effectiveness is thoroughly assessed in contrast to other methodology and approaches.

Judgment Prediction from Legal Documents using Texas Wolf Optimization-based Deep BiLSTM

The Texas Wolf Optimisation (TWO) algorithm is utilised to optimise a Deep Bidirectional Long Short-Term Memory (BiLSTM) model for the purpose of predicting judgements from legal documents. The TWO technique enhances the model's ability to grasp complex patterns in legal texts by optimising hyperparameters. The BiLSTM architecture employs bidirectional processing to understand the context of text by taking into account both preceding and subsequent scenarios. This process begins by extracting features from legal texts, which are subsequently entered into the optimised BiLSTM network. The result is an extremely accurate model for predicting judicial outcomes, which considerably aids legal practitioners in making decisions and improves overall performance.

Experimental Set Up

The judicial judgement prediction experiment is conducted using the Python programming language on a computer system equipped with 8GB of RAM and the Windows 10 operating system.

Dataset Description

A legal judgement prediction database in the context of the Supreme Court is a well-organised compilation of past legal cases, namely those that have been resolved by the highest court in a certain nation. The document provides comprehensive data on previous instances, encompassing the involved parties, legal matters, and ultimate verdicts. The system incorporates case-specific characteristics such as citations, statutes, and precedents, enabling algorithms to detect trends and factors that influence judicial rulings. Legal analytics is significantly enhanced and the accuracy of legal judgement predictions is improved, resulting in substantial benefits for the legal profession and the justice system.

Performance analysis based on TP

Figure 7 displays the performance metrics of the TWO-BiLSTM models in predicting court judgements. Figure 4a demonstrates the accuracy values of the TWO-BiLSTM technique, which are 91.55%, 93.59%, 95.62%, 96.03%, and 97.00% for epochs 100, 200, 300, 400, and 500, respectively. Throughout these epochs, the Training Percentage (TP) remains continuously at 90.

Referring to Figure 4b, the TWO-BiLSTM approach achieves impressive f-score outcomes of 91.58%, 93.64%, 95.60%, 96.04%, and 97.29%, while maintaining a true positive (TP) count

of 90.

The TWO-BiLSTM method achieves precision rates of 91.26%, 93.01%, 95.71%, 96.00%, and 97.10% in Figure 4c, while consistently maintaining a true positive (TP) value of 90. Figure 4d displays the recall values for the TWO-BiLSTM approach. The approach achieves recall values of 91.90%, 94.28%, 95.21%, 9.79%, and 97.19% across epochs. The approach also maintains a consistent true positive (TP) value of 90. The results emphasise the strong and consistent performance of the TWO-BiLSTM model in predicting judicial judgements at different epochs, indicating its potential as a dependable tool for this purpose.

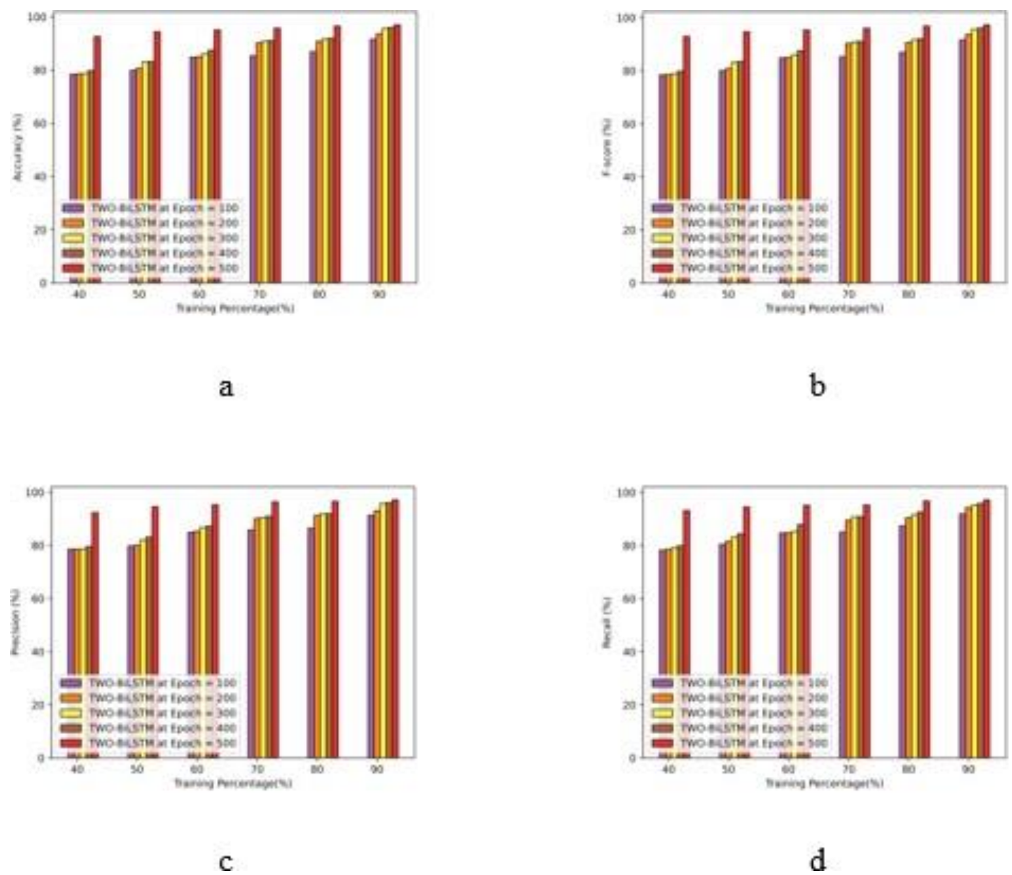


Figure 7 Performance analysis based on TP a) accuracy, b) f-score, c) precision and d) recall

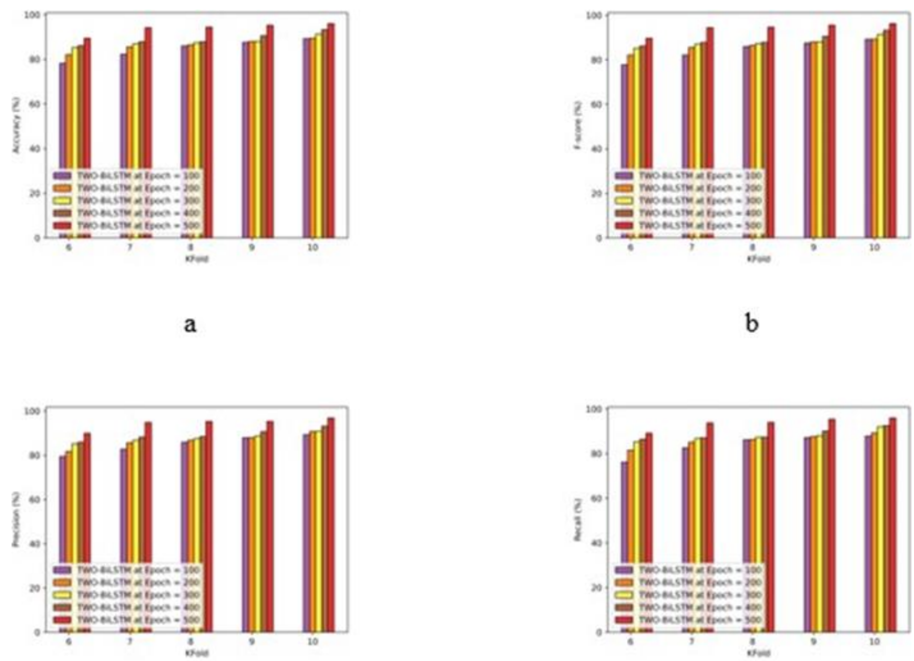
Performance analysis based on K-fold

The performance metrics of the TWO-BiLSTM models in legal judgement prediction are shown in Figure 8. Figure 8a demonstrates the accuracy values of the TWO-BiLSTM technique, which are 91.55%, 93.59%, 95.62%, 96.03%, and 97.00% for epochs 100, 200, 300, 400, and 500, respectively. The approach constantly maintains a K-fold 10. Referring to Figure 8b, the TWO-BiLSTM approach demonstrates impressive f-score outcomes of 89.23%, 89.25%, 91.43%, 93.27%, and 96.25%, while maintaining a k-fold value

of 10.

In Figure 8c, the TWO-BiLSTM method achieves precision values of 89.27%, 90.74%, 90.84%, 93.14%, and 96.89%, while consistently maintaining a k-fold 10. Figure 8sd displays the recall values for the TWO-BiLSTM technique. The approach achieves recall values of 87.82%, 89.20%, 92.06%, 92.50%, and 95.96% across epochs. Additionally, it maintains a consistent true positive (TP) value of 90.

The results emphasise the strong and consistent performance of the TWO-BiLSTM model in predicting judicial judgements at different epochs, indicating its potential as a dependable tool for this purpose.



a b

c d

Figure 8 Performance analysis based on K-fold a) accuracy, b) f-score, c) precision and d) recall

Comparative Methods

An evaluative study was performed to determine the effectiveness of the TWO-BiLSTM model in predicting court judgements. This evaluation encompassed an extensive examination of diverse techniques and models, such as pre-trained embeddings, the TenLa algorithm, the hierarchical attention neural network, LSTM and CNN, BiLSTM, as well as two innovative approaches, HHO-BiLSTM and GWO-BiLSTM, which were introduced in this study. The purpose of this comprehensive comparison was to emphasise the unique benefits and exceptional performance of the TWO-BiLSTM model in this particular application.

Comparative analysis based on TP

Figure 9a demonstrates the precision of the TWO-BiLSTM model in forecasting court judgements, with a True Positive (TP) rate of 90. The TWO-BiLSTM model demonstrates superior performance compared to the GWO-BiLSTM model, with a significant margin of 10.07%, resulting in an accuracy of 97.00%.

Now let's examine Figure 9b, which focusses on the f-score of the TWO-BiLSTM model for predicting judicial judgements, while maintaining a true positive rate of 90%. The TWO-

BiLSTM model outperforms the GWO-BiLSTM model, achieving an impressive f-score of 97.29%, exceeding its counterpart by a significant margin of 10.00%.

Figure 9c demonstrates the accuracy gained by the TWO-BiLSTM model in predicting court judgements, with a constant true positive rate of 90%. The TWO-BiLSTM model achieves a precision of 97.10%, surpassing the GWO-BiLSTM model by a significant margin of 10.42%.

Finally, in Figure 9d, display the recall attained by the TWO-BiLSTM model for legal judgement prediction, while keeping the true positive rate (TP) at 90%. The TWO-BiLSTM model demonstrates superior performance compared to the GWO-BiLSTM model, with a notable improvement of 9.58%. It achieves a recall rate of 97.19%.

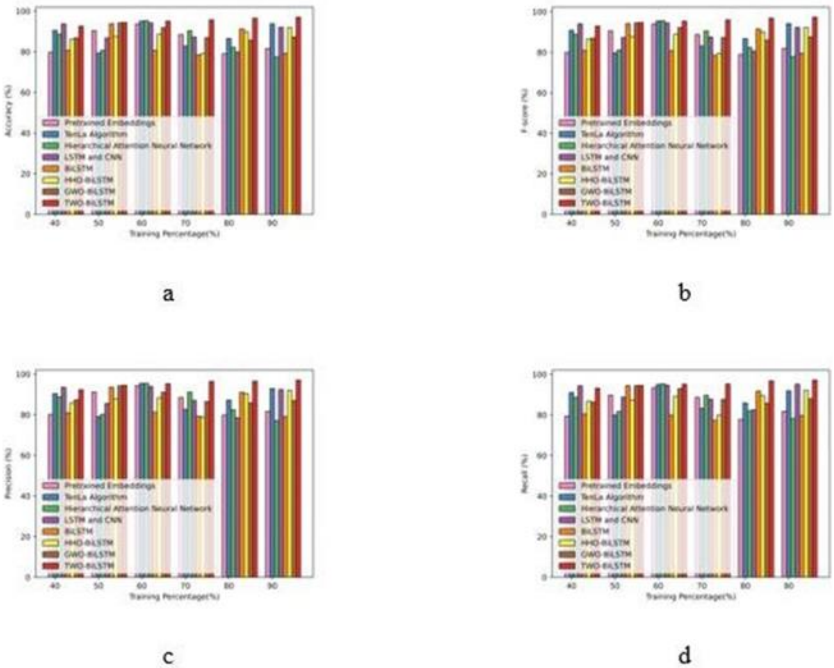


Figure 9 Comparative analysis based on TP a) accuracy, b) f-score, c) precision and d) recall

Comparative analysis based on K-fold

Figure 10a demonstrates the precision attained by the TWO-BiLSTM model in forecasting legal rulings, while utilising a K-fold 10. The TWO-BiLSTM model demonstrates superior performance compared to the GWO-BiLSTM model, with a significant margin of 7.18%, resulting in an accuracy of 96.00%.

Now let's examine Figure 10b, which focusses on the f-score of the TWO-BiLSTM model for predicting judicial judgements. We will once again maintain a K-fold value of 10. The TWO- BiLSTM model outperforms the GWO-BiLSTM model, achieving an impressive f-score of 96.25%, exceeding its counterpart by a significant margin of 6.98%.

Figure 10c demonstrates the accuracy attained by the TWO-BiLSTM model in predicting court judgements, using a 10-fold cross-validation. The TWO-BiLSTM model achieves a precision of 96.89%, surpassing the GWO-BiLSTM model by a significant margin of 9.14%. Finally, in Figure 10d, display the recall attained by the TWO-BiLSTM model for legal judgement prediction, while still using K-fold 10. The TWO-BiLSTM model demonstrates superior performance compared to the GWO-BiLSTM model, with an impressive improvement of 95.96% and achieving a recall rate of 5.37%.

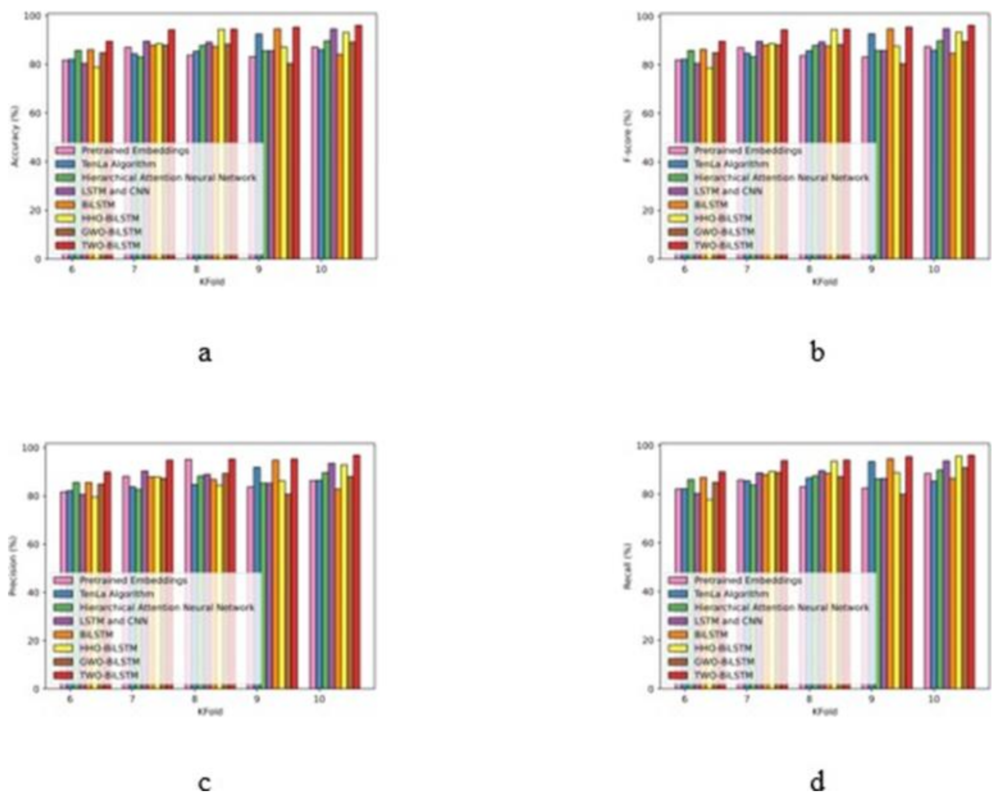


Figure 10 Comparative analysis based on K-fold a) accuracy, b) f-score, c) precision and d) recall

Comparative Discussion

The comparative analysis was conducted on a database with a true positive (TP) value set at 90 and a k-fold value of 10. The objective was to demonstrate the exceptional superiority of the TWO-BiLSTM models in comparison to the currently available models. The performance measures, measured at the 90th percentile, showed exceptional effectiveness, with values of 97.00, 97.29, 97.10, and 97.19. In the specific scenario of k-fold 10, the TWO-BiLSTM models consistently demonstrated their exceptional performance, obtaining remarkable metrics with values of 96.00, 96.25, 96.89, and 95.96. This thorough assessment highlights the significant progress that the TWO-BiLSTM models offer in contrast to their predecessors.

Table 4. Comparative discussion table for TP and K-fold

Models	**TP 90**				**K-fold 10**			
	Accur acy	**F-score**	**Precision**	**Recall**	**Accuracy**	**F-score**	**Precision**	**Recall**
Pre trained Embeddings	81.62	81.87	81.71	81.79	87.12	87.54	86.38	88.47
Ten La Algorithm	93.79	94.14	92.88	91.86	85.95	86.01	86.50	85.30
Hierarchical AttentionNeural Network	77.41	77.80	77.11	78.27	89.63	89.91	89.68	89.87
LSTM and CNN	92.03	92.2 8	92.42	95.14	94.49	94.85	93.52	93.62

BiLSTM	79.17	79.48	79.04	79.68	84.08	84.73	82.83	86.47
HHO-BiLSTM	91.91	92.19	92.01	92.10	93.08	93.39	92.89	95.50
GWO-BiLSTM	87.24	87.56	86.98	87.88	89.11	89.54	88.04	90.81
Proposed TWO-BiLSTM	97.00	97.29	97.10	97.19	96.00	96.25	96.89	95.96

Table 4 displays a comparative examination of various models utilising True Positive (TP) and K-fold cross-validation metrics. The evaluated models consist of Pre-trained Embeddings, Ten La Algorithm, Hierarchical Attention Neural Network, LSTM & CNN, BiLSTM, HHO- BiLSTM, GWO-BiLSTM, and the Proposed TWO-BiLSTM. The table presents a comparison of the accuracy, F-score, precision, and recall of each model under two conditions: TP 90 and K-fold 10. The Proposed TWO-BiLSTM model demonstrates superior performance compared to previous models, with an impressive accuracy of 97% and an F-score of 97.29% in TP 90. It also exhibits great performance in K-fold 10, with an accuracy of 96%. The superior effectiveness of the Proposed TWO-BiLSTM model is seen in both metrics.

CONCLUSION

Overall, this work effectively shows that the Texas Wolf Optimisation (TWO) algorithm improves the performance of a Deep Bidirectional Long Short-Term Memory (BiLSTM) model for predicting judicial judgements. Through the optimisation of hyperparameters, the TWO-BiLSTM model attains higher levels of accuracy, precision, recall, and F1-scores in comparison to conventional models and other sophisticated techniques such as GWO-BiLSTM and HHO-BiLSTM. The model's bidirectional processing enables it to catch complex patterns in legal texts, leading to a strong prediction system that outperforms other similar systems. By conducting a thorough assessment, which involved analysing performance using True Positive (TP) rates and K-fold cross-validation, the TWO-BiLSTM model repeatedly shown its capacity to accurately forecast judicial outcomes. This emphasises its capacity as a dependable instrument for legal professionals, facilitating decision-making and contributing to the progress of legal analytics. The comparison analysis highlights the significant enhancements provided by the TWO-BiLSTM model, establishing its value as a beneficial addition to the field of judicial judgement prediction. Subsequent investigations could examine the incorporation of supplementary optimisation methods and the utilisation of this methodology in other legal fields to augment its overall applicability and efficacy.

REFERENCES

J. Zhu, J. Wu, X. Luo, and J. Liu, “Semantic matching based legal information retrieval,” Artif. Intell. Law, vol. 32, no. 2, pp. 397–426, 2024, doi: 10.1007/s10506-023-09354-x.

J. J. Nay, “LAW INFORMS CODE : A LEGAL,” vol. 20, no. 3, 2023.

A. Modi et al., “SemEval 2023 Task 6: LegalEval - Understanding Legal Texts,” 2023.

Z. Fei et al., “LawBench: Benchmarking Legal Knowledge of Large Language Models,” pp. 1–38, 2023.

J. Cui, X. Shen, and S. Wen, “A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges,” IEEE Access, vol. 11, no. September, pp. 102050– 102071, 2023, doi: 10.1109/ACCESS.2023.3317083.

A. Burnap and J. R. Hauser, “Product Aesthetic Design : A Machine Learning Augmentation,” vol. 42, no. 6, pp. 1029–1056, 2023, doi: 10.1287/mksc.2022.1429.

A. Kale, T. Nguyen, F. C. H. Jr, C. Li, J. Zhang, and X. Ma, “Provenance documentation to enable explainable and trustworthy AI : A literature review,” pp. 0–2, 2023, doi: 10.1162/dint.

R. Cotterell and Z. Eth, “On the Role of Negative Precedent in Legal Outcome Prediction,” vol. 11, pp. 34–48, 2023.

Y. Lyu et al., “Improving legal judgment prediction through reinforced criminal element extraction,” Inf. Process. Manag., vol. 59, no. 1, p. 102780, 2022, doi: 10.1016/j.ipm.2021.102780.

A. Ieven, M. F. Moens, T. Gonc, and P. Quaresma, “ECHR : Legal Corpus for Argument Mining,” pp. 67–75, 2020.

P. H. L. De Araujo, T. E. De Campos, F. A. Braz, and N. C. Silva, “VICTOR : a dataset for Brazilian legal documents classification,” no. May, pp. 1449–1458, 2020.

P. Kalamkar et al., "Named Entity Recognition in Indian court judgments," 2020.

K. Atkinson, T. Bench-capon, and D. Bollegala, "Explanation in AI and law : Past , present and future ☆," Artif. Intell., vol. 289, no. November 2019, p. 103387, 2020, doi: 10.1016/j.artint.2020.103387.

B. Heinrichs and S. B. Eickhoff, "Your evidence ? Machine learning algorithms for medical diagnosis and prediction," no. November 2019, pp. 1435–1444, 2020, doi: 10.1002/hbm.24886.

Y. Huang, W. Dai, J. Yang, and L. Cai, "Semi-Supervised Abductive Learning and Its Application to Theft Judicial Sentencing," no. Icdm, pp. 1070–1075, 2020, doi: 10.1109/ICDM50108.2020.00127.

C. V Trappey, A. J. C. Trappey, and B. Liu, "Identify trademark legal case precedents - Using machine learning to enable semantic analysis of judgments," vol. 62, no. June 2019, 2020.

F. Contini, "Artificial Intelligence and the Transformation of Humans , Law and Technology Interactions in Judicial Proceedings," vol. 2, no. 1, pp. 4–18, 2020.

D. Trautmann, A. Petrova, and F. Schilder, "Legal Prompt Engineering for Multilingual Legal Judgement Prediction," 2022.

D. U. Shengyong, "An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning," IEEE Access, vol. 10, pp. 22516–22532, 2022, doi: 10.1109/ACCESS.2022.3153478.

O. Sen et al., "Bangla Natural Language Processing : A Comprehensive Analysis of Classical , Machine Learning , and Deep Learning-Based Methods," IEEE Access, vol. 10, pp. 38999–39044, 2022, doi: 10.1109/ACCESS.2022.3165563.

D. Licari and G. Comandè, "ITALIAN-LEGAL-BERT : A Pre-trained Transformer Language Model for Italian Law," 2022.

W. Hwang, D. Lee, K. Cho, H. Lee, and M. Seo, "A Multi-Task Benchmark for Korean
Legal Language Understanding and Judgement Prediction," Adv. Neural Inf. Process. Syst., vol. 35, no. NeurIPS, pp. 1–15, 2022.

S. Greenstein, Preserving the rule of law in the era of artificial intelligence (AI), vol. 30, no. 3. Springer Netherlands, 2022. doi: 10.1007/s10506-021-09294-4.

A. Tagarelli and A. Simeri, Unsupervised law article mining based on deep pre - trained language representation models with application to the Italian civil code, vol. 30, no. 3. Springer Netherlands, 2022. doi: 10.1007/s10506-021-09301-8.

J. Dhanani, R. Mehta, and D. Rana, "Effective and scalable legal judgment recommendation using pre-learned word embedding," Complex Intell. Syst., vol. 8, no. 4, pp. 3199–3213, 2022, doi: 10.1007/s40747-022-00673-1.

D. Biesner et al., "Anonymization of German financial documents using neural network- based language models with contextual word representations," Int. J. Data Sci. Anal., vol. 13, no. 2, pp. 151–161, 2022, doi: 10.1007/s41060-021-00285-x.

L. B. Coelho, "Reviewing machine learning of corrosion prediction in a data-oriented perspective," 2022, doi: 10.1038/s41529-022-00218-4.

W. E. I. E. Zhang, M. Guo, H. U. Wang, and Q. Z. Sheng, "Multi-document Summarization via Deep Learning arXiv : 2011 . 04843v3 [cs . CL] 9 Dec 2021," vol. 1, no. 1, pp. 1–35, 2021.

U. C. Berkeley, C. Burns, and S. Ball, "CUAD : An Expert-Annotated NLP Dataset for," no. NeurIPS, 2021.

V. Malik et al., "ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation," ACL-IJCNLP 2021 - 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 4046–4062, 2021, doi: 10.18653/v1/2021.acl-long.313.

V. Malik, "Semantic Segmentation of Legal Documents via Rhetorical Roles," 2021.

Q. Dong and S. Niu, "Legal Judgment Prediction via Relational Learning," pp. 983– 992, 2021, doi: 10.1145/3404835.3462931.

S. Wachter, B. Mittelstadt, C. Russell, and N. Law, "Bias Preservation in Machine Learning : The Legality of Fairness Metrics Under EU Non-Discrimination Law," vol. 123, no. 3, 2021.

A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis) contents : A survey of dataset development and use in machine learning research," Patterns, vol. 2, no. 11, p. 100336, 2021, doi: 10.1016/j.patter.2021.100336.

F. Dadgostari, M. Guim, P. A. Beling, M. A. Livermore, and D. N. Rockmore, Modeling law search as prediction, vol. 29, no. 1. Springer Netherlands, 2021. doi: 10.1007/s10506-020-09261-5.

N. Shah, N. Bhagat, and M. Shah, "Crime forecasting : a machine learning and computer vision approach to crime prediction and prevention," 2021.

C. I. Hausladen, M. H. Schubert, and E. Ash, "Text classification of ideological direction in judicial opinions," Int. Rev. Law Econ., vol. 62, pp. 0–19, 2020, doi: 10.1016/j.irle.2020.105903.

A. Deeks, "High-tech international law," George Washington Law Rev., vol. 88, no. 3, pp. 574–653, 2020.

J. Valvoda, R. Cotterell, and S. Teufel, "On the Role of Negative Precedent in Legal Outcome Prediction," Trans. Assoc. Comput. Linguist., vol. 11, pp. 34–48, 2023, doi: 10.1162/tacl_a_00532.

P. Kalamkar, A. Agarwal, A. Tiwari, S. Gupta, S. Karn, and V. Raghavan, "Named Entity Recognition in Indian court judgments," NLLP 2022 - Nat. Leg. Lang. Process. Work. 2022, Proc. Work., pp. 184–193, 2022, doi: 10.18653/v1/2022.nllp-1.15.

Y. Feng, C. Li, and V. Ng, "Legal Judgment Prediction: A Survey of the State of the
Art," IJCAI Int. Jt. Conf. Artif. Intell., pp. 5461–5469, 2022, doi: 10.24963/ijcai.2022/765.