Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 5, pp. 3822–3831, DOI 10.12694/scpe.v25i5.3002

# COPYRIGHT PROTECTION AND RISK ASSESSMENT BASED ON INFORMATION EXTRACTION AND MACHINE LEARNING: THE CASE OF ONLINE LITERARY WORKS

XUDONG LIN*

**Abstract.** With the proliferation of digital platforms, the dissemination of literary works has encountered unprecedented challenges, particularly concerning copyright infringement and unauthorized use. This study introduces a comprehensive framework for copyright protection and risk assessment, specifically tailored to online literary works. The framework employs advanced CNN based information extraction (IE) techniques coupled with machine learning (ML) algorithms to identify, classify, and protect literary content against copyright violations. Firstly, we delineate a novel CNN-Decision tree-based IE methodology that systematically harvests metadata and textual content from various online repositories. This process is designed to detect and index online literary works, extracting pertinent features such as authorship, publication date, and textual patterns. Following the extraction, the study utilizes natural language processing (NLP) to analyze and compare content, pinpointing potential instances of copyright infringement by identifying significant overlaps and stylistic similarities with registered works. Subsequently, we introduce a risk assessment model developed through supervised machine learning. This model is trained on a labelled dataset comprising instances of both copyrighted and non-copyrighted works, along with known cases of copyright infringement. By analyzing the extracted features, the model assesses the probability of infringement, categorizing risks into high, medium, and low categories. This stratification allows stakeholders to prioritize enforcement actions and resources efficiently. The study further explores the implementation of various ML algorithms, including decision trees, support vector machines, and neural networks, to determine the most effective approach for copyright protection in the literary domain. We evaluate the models based on accuracy, precision, recall, and F1-score metrics, emphasizing their capacity to generalize and operate in dynamic, real-world environments.

**Key words:** information extraction, Copyright Protection, risk assessment.

**1. Introduction.** In the age of digital media, the protection of intellectual property has emerged as a paramount concern, particularly within the creative industries. Copyright laws serve as the bulwark against unauthorized use and reproduction of original works, safeguarding the interests and rights of creators and ensuring that they receive recognition and economic benefits from their contributions. However, as the digital footprint of society expands, copyright protection confronts increasingly complex challenges that necessitate advanced research and innovation. The pertinence of copyright protection is multifaceted. It not only upholds the moral and legal rights of authors but also fosters a thriving ecosystem for cultural and creative growth. By ensuring creators can benefit from their works, copyright stimulates investment in creativity and innovation, driving the growth of industries ranging from publishing to entertainment. Yet, the rapid evolution of technology has outpaced the traditional mechanisms of copyright enforcement, making it imperative to explore new avenues that can adequately respond to the scale and sophistication of copyright infringement in the digital realm.

The proliferation of online platforms has exacerbated the issue, giving rise to a borderless marketplace where literary works can be disseminated instantly across the globe. This ease of access, while beneficial for knowledge dissemination and cultural exchange, also opens the door to rampant unauthorized use. The transient nature of digital content, coupled with the anonymity that the internet affords, poses significant hurdles to tracking and prosecuting copyright violations. Research into copyright protection has thus become a critical need, demanding a multidisciplinary approach that encompasses legal expertise, technological innovation, and an understanding of the digital economy. Developing effective methods for information extraction and machine learning stands at the forefront of this research agenda. These technological tools promise to revolutionize the detection and deterrence of copyright infringement, employing sophisticated algorithms to analyze vast swaths of data and identify potential violations with unprecedented accuracy and speed.

*Educational and Scientific Institute of International Relations,Taras Shevchenko National University of Kyiv, Kyiv, Ukraine 01033 (xudonglinst@outlook.com)

Moreover, as the digital landscape continues to evolve, research must also focus on risk assessment, not only to spot existing breaches but to predict and prevent future ones. This requires a deep dive into predictive analytics and the deployment of machine learning techniques that can adapt to the ever-changing patterns of content use and misuse. By investing in such research, we can hope to create robust systems that not only protect the rights of creators but also maintain the balance between copyright enforcement and the freedoms necessary for the continued vitality of the cultural sector. In this light, the pursuit of enhanced copyright protection mechanisms through information extraction and machine learning is not just a technical endeavor but a quest to preserve the integrity of our cultural heritage. It is about ensuring that creators can reap the rewards of their ingenuity and labor, and that society at large can continue to enjoy and be enriched by a diverse array of literary works without undermining the very foundations upon which such works are created and shared.

The main contribution of the article is,

1. The study develops a new information extraction (IE) approach that integrates Convolutional Neural Networks (CNNs) with Decision Trees to harvest metadata and textual content from various online literary repositories.
2. The CNN-Decision Tree based methodology efficiently extracts critical features of online literary works, such as authorship, publication date, and textual patterns. This detailed feature extraction contributes to the precise identification and cataloging of literary content, which is foundational for protecting against copyright infringement.
3. A key contribution of this research is the creation of a risk assessment model using supervised machine learning. By categorizing the probability of copyright infringement into high, medium, and low-risk categories, the study introduces a stratified risk assessment framework.

**2. Related work.** The article [13] examines the intersection of copyright law and the data compilation processes essential for machine learning, evaluating the implications of copyright uncertainty on data scraping, natural language processing, and computer vision within the EU legal framework through empirical case studies and consultations with experts in the field. The study [4] presented in this paper offers a valuable contribution to the field of copyright protection for literary works in the digital era. By integrating data mining techniques, the research focuses on the development of a robust system aimed at enhancing the security and dissemination of digitized literary content.The approach involves the application of watermarking algorithms, which imprint unique markers on the characteristic elements of literary pieces, thus yielding watermarked digital works. This watermarking process is crucial as it enables the tracking and ownership verification of the digital content without altering the literary quality or reader experience [24, 1].

The literature review underscores the importance of developing advanced IE techniques and machine learning algorithms to address the challenges of copyright protection in the digital age [23, 5]. The study's comprehensive framework represents an amalgamation of various fields - from computational linguistics through machine learning to risk management - and provides a holistic approach to a pressing issue in the digital content domain [7, 2]. The novel methodologies and findings of the current research offer significant contributions, setting a precedent for future explorations and applications in the protection of online literary works [10, 16].

**3. Proposed methodology.** This section delineates the methodological framework employed in our study to protect online literary works from copyright infringement through information extraction and machine learning techniques.

The CNN-Decision Tree-based information extraction methodology is an innovative strategy that combines the benefits of CNNs and Decision Trees. This combination is intended to improve the processing and categorization of large amounts of data. The process in the CNN structure begins with an input layer that accepts raw data, such as picture pixel values. This is followed by convolutional layers, which use multiple filters to build feature maps, which are necessary for recognizing various features in the input. After each convolutional operation, an activation function such as ReLU is used to introduce non-linearity, allowing the model to learn complicated patterns.

Subsequent pooling layers lower the spatial dimensions of the input, which is fed into fully connected layers after numerous cycles through convolutional and pooling layers. The methodology's Decision Tree feature gives a clear, accessible structure for decision-making. Decision Trees are tree-like models in which each internal node

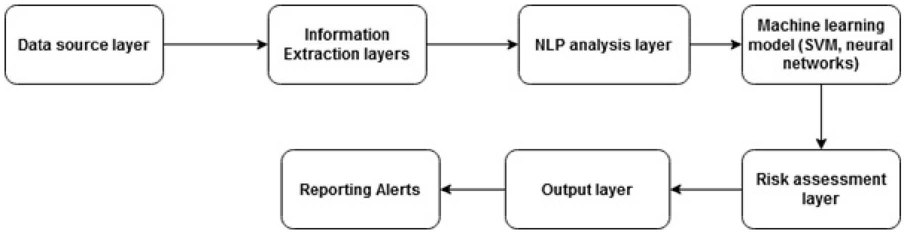


Fig. 3.1: Proposed Copyright Risk Assessment Architecture using ML

Table 3.1: Sample annotated data

| **Entity type** | **Counts** | **Example** |
|---|---|---|
| PER | 9,383 | my mother, Jarndyce, the doctor, a fool, his companion |
| FAC | 2,154 | the house, the room, the garden, the drawing-room, the library |
| LOC | 1,170 | the sea, the river, the country, the woods, the forest |
| GPE | 878 | London, England, the town, New York, the village |
| VEH | 197 | the ship, the car, the train, the boat, the carriage |

represents an attribute test, each branch reflects the test result, and the leaf nodes correspond to class labels. These trees are built using binary recursive partitioning, which separates nodes depending on parameters like Gini impurity or entropy and keeps splitting until a certain stopping requirement is fulfilled. This might be the tree's present depth or another parameter. Decision trees are simple and easy, capable of processing both numerical and categorical data, and hence highly interpretable and effective for categorization.

Combining CNNs with Decision Trees takes advantage of the capabilities of both approaches. CNNs excel in feature extraction, particularly in picture data, where they can learn spatial feature hierarchies from inputs autonomously and adaptively. Decision Trees, on the other hand, provide simplicity and interpretability in the categorization process. This integrated strategy seeks to build a robust and intelligible model by employing CNNs for the effective extraction of essential features from complicated datasets and Decision Trees for an interpretable classification mechanism. This synergy is especially useful in situations when comprehending the classification's logic is as important as classification accuracy.

**3.1. Dataset.** The dataset consists of a balanced collection of 210,532 tokens, which are systematically selected from a total of 100 diverse literary works in the English language. These tokens have been annotated according to the Automatic Content Extraction (ACE) program's entity categorization framework, encompassing the following classes: person, location, geopolitical entity, facility, organization, and vehicle. This is publicaly available dataset on link https://github.com/dbamman/litbank. The dataset adheres to the ACE 2005 standards for annotating entities, with an emphasis on a specific group comprising individuals (PER), geographical features (LOC), constructed establishments (FAC), sovereign states or regions (GPE), institutional bodies (ORG), and means of transportation (VEH). Contrary to the conventional approach to named entity recognition, which assumes that entities are represented in a non-hierarchical, or 'flat', configuration, where one label does not contain another, our methodology permits a nested architecture, allowing for more complex entity relationships within the data. The table 3.1 shows sample dataset annotation details.

**3.2. Information Extraction Methodology.** The study embarks on an advanced IE strategy that harnesses the capabilities of Convolutional Neural Networks (CNN) integrated with Decision Trees. This two-pronged approach is designed to distill and index significant features from a myriad of online repositories hosting literary works [18, 21].

Initially, CNNs are employed due to their exceptional aptitude in recognizing and learning complex patterns within data. For textual content analysis, a bespoke CNN architecture is adopted, featuring convolutional layers tailored to discern linguistic patterns, semantic structures, and stylometric features that are indicative of authorship and originality [17, 12]. Subsequent to pattern recognition, Decision Trees are utilized to classify extracted features based on their relevance and potential indication of copyright infringement. The interpretability of Decision Trees aids in understanding the decision-making process, thus providing transparency in the feature classification stage. Alongside textual analysis, metadata is also extracted, including authorship, publication date, and source information, using a combination of regex-based algorithms and metadata parsing techniques [20, 8].

**3.3. Convolutional Neural Networks (CNN) for Feature Learning.** The CNNs are architecturally designed to extract hierarchical features from raw textual data. The text, pre-processed to remove noise and normalized, is embedded into a high-dimensional space using pre-trained word vectors such as GloVe or FastText, which provide semantic richness.

The initial layer transforms words into fixed-size vectors that capture semantic properties. Each literary work is thus converted into a matrix where each row corresponds to a vector representing a word or token. Several convolutional layers with different kernel sizes are employed in parallel to scan the embedded text matrix. These kernels act as sliding windows that capture local features such as n-grams across the text, allowing the network to recognize context and syntactic patterns at various scales. Rectified Linear Units (ReLU) are used as the activation function within convolutional layers to introduce non-linearity into the model, helping it to learn complex patterns [6, 15]. Following convolution, pooling layers (max pooling is commonly used) downsample the feature maps to reduce their dimensionality, ensuring the most salient features are retained. This step reduces computation and mitigates the risk of overfitting. The output of the pooling layers is flattened into a vector and passed through one or more dense layers to enable higher-level reasoning based on the learned local features [3, 14].

**3.4. Decision Trees for Feature Classification.** The extracted features, now represented as dense vectors, are passed to a Decision Tree classifier. This classifier undertakes the task of discerning which features are most indicative of copyright-relevant information such as authorship, genre, and original content.

Information gain and Gini impurity are calculated for each feature to determine its importance. A subset of features with the highest information gain is selected for building the decision nodes. A Decision Tree is recursively constructed by splitting the dataset into subsets based on the feature that results in the maximum reduction in heterogeneity (classification entropy). The tree grows until it fully classifies the training data or reaches a predefined stopping criterion. To avoid overfitting, the tree is pruned back. Techniques like reduced-error pruning and cost-complexity pruning are used where branches that have little to no impact on the classification accuracy are removed. Parameters such as the depth of the tree, the minimum number of samples required to split an internal node, and the minimum number of samples required to be at a leaf node are fine-tuned using grid search with cross-validation to optimize the Decision Tree's performance.

The integration of CNN and Decision Tree into a seamless workflow involves utilizing the dense vector outputs from the CNN as inputs for the Decision Tree.

*Combining Outputs.* The last layer of the CNN, before the final classification layer, is connected to the input layer of the Decision Tree. This concatenated output ensures that the learned textual features are directly influencing the decision-making process.

*Ensemble Learning.* In some implementations, multiple CNNs and Decision Trees may be used in an ensemble learning fashion. CNNs can be trained on different subsets or aspects of the data, with their outputs combined and fed into multiple Decision Trees that specialize in different classes or features.

*Model Evaluation.* The hybrid model is evaluated using a hold-out validation set. Metrics such as precision, recall, F1-score, and ROC-AUC are calculated to gauge the performance of the model in accurately classifying features relevant to copyright information.

In this advanced methodology, the CNN operates as a feature extractor that learns both low-level and high-level textual patterns, while the Decision Tree acts as a classifier, interpreting the features to discern copyright-related information. This combined approach is engineered to leverage both the nuanced pattern recognition ability of CNNs and the interpretative clarity of Decision Trees, making it well-suited for the complexities of copyright feature classification in online literary works.

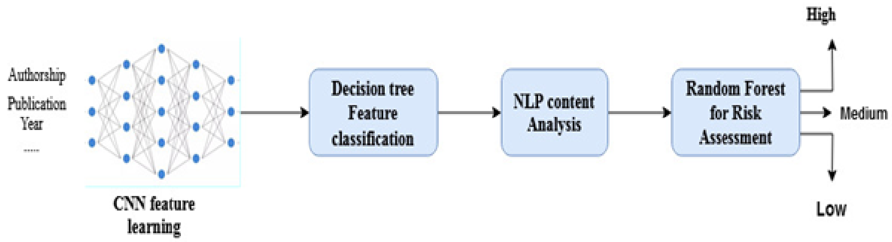


Fig. 3.2: Proposed Risk Assessment Model

**3.5. Natural Language Processing (NLP) for Content Analysis.** Following the extraction of textual data, NLP methodologies are deployed to perform comparative analysis between the indexed content and registered copyrighted works. Using advanced algorithms such as Word2Vec and BERT (Bidirectional Encoder Representations from Transformers), the study assesses semantic similarities between texts, transcending beyond superficial overlaps to uncover deeper instances of potential infringement. Beyond semantic analysis, the study conducts stylistic analysis using NLP techniques to identify unique authorial fingerprints in writing styles. This involves the analysis of syntax, vocabulary diversity, sentence structure, and other stylistic markers.

**3.6. Machine Learning for Copyright Risk Assessment.** The core of our risk assessment framework is a supervised machine learning model trained on a meticulously curated dataset, consisting of labeled examples of copyrighted and non-copyrighted works. The dataset is divided into training, validation, and test sets. Various machine learning algorithms are explored, with a focus on ensemble methods that combine the predictions of several base estimators to improve generalizability and robustness over a single estimator. Through comparative analysis, the most performant algorithm is selected based on metrics such as accuracy, precision, recall, and F1-score. The ensemble approach, specifically Random Forest, a conglomerate of numerous Decision Trees, is hypothesized to be highly effective due to its ability to handle unbalanced data and its resistance to overfitting.

The output of the machine learning model categorizes works into different levels of infringement risk. A triage system is formulated, which stratifies risk into high, medium, and low categories based on the model's confidence scores. This triage system allows for prioritized response actions. Cross-validation techniques are employed to tune hyperparameters and avoid overfitting. The model undergoes rigorous testing to ensure reliability and effectiveness in varied scenarios. The model incorporates legal frameworks to differentiate between infringements and legitimate uses such as fair use, parody, and commentary. Ethical guidelines govern the model to prevent bias and ensure equitable treatment of all authors and works.

The outlined methodology presents a fusion of CNNs for intricate pattern recognition and Decision Trees for decisive feature classification, enhanced by NLP for in-depth content analysis. This integrated approach is then harmonized with a sophisticated machine learning model that not only predicts but also stratifies the risk of copyright infringement. Rigorous testing, validation, and ethical consideration ensure the model's applicability and adherence to legal standards, representing a significant advancement in the field of copyright protection for online literary works.

## 4. Result analysis.

**4.1. Result evaluation.** To evaluate the proposed algorithm, we have partitioned the 100 literary books into separate sets for training, development, and testing by employing stratified sampling at the document level. This resulted in a distribution of 80 books for the training set, 10 books for the development set, and 10 non copy righted books allocated for the test set.

Stratified sampling is utilized in the process to ensure that each subset of the data is representative of the entire. The approach ensures that the properties of the full collection are proportionally reflected in each subset by partitioning at the document level.

This implies that each set (training, development, and testing) has a mix of different literary styles, times, and genres, preserving the original dataset's richness and complexity. This large chunk, consisting of 80 books,

**Algorithm 1** Copyright Protection Model

```
Input: Dataset of literary works with features and copyright status labels
Output: Risk assessment categorizing works into high, medium, and low infringement risk
Begin:                                                        ▷ Preprocessing Textual Data
procedure PREPROCESS_TEXT(Data)
    for each literary_work in Data do
        Clean and normalize the text
        Tokenize the text into words or characters
        Embed the tokens using pre-trained word vectors (e.g., GloVe, FastText)
    end for
end procedure                                                 ▷ CNN for Feature Learning
procedure TRAIN_CNN(Text_Embeddings)
    Initialize CNN with convolutional layers, ReLU activations, and max pooling
    for each epoch do
        for each batch in Text_Embeddings do
            Perform forward propagation through CNN layers
            Apply backpropagation and update CNN weights
        end for
    end for
end procedure                                                 ▷ Decision Trees for Feature Classification
procedure TRAIN_DECISION_TREE(Features)
    Initialize Decision Tree with entropy or Gini impurity criteria
    for each feature_vector in Features do
        Calculate information gain for each feature
        Build decision tree based on maximum information gain
        Prune the tree to avoid overfitting
    end for
end procedure                                                 ▷ NLP for Content Analysis
procedure PERFORM_CONTENT_ANALYSIS(Indexed_Content, Copyrighted_Works)
    for each content_pair in (Indexed_Content, Copyrighted_Works) do
        Analyze semantic and stylistic similarities
        Use NLP algorithms like Word2Vec and BERT for deep analysis
    end for
end procedure                                                 ▷ Machine Learning for Risk Assessment
procedure TRAIN_RISK_ASSESSMENT_MODEL(Labeled_Dataset)
    Split Labeled_Dataset into training, validation, and test sets
    Explore various machine learning algorithms, including ensemble methods
    Select the best-performing algorithm based on validation metrics
    Train the final model on the training set
    Evaluate model performance on the test set using precision, recall, F1-score
end procedure                                                 ▷ Main Program
Dataset = Load all literary works data
Text_Embeddings = PREPROCESS_TEXT(Dataset)
CNN_Features = TRAIN_CNN(Text_Embeddings)
Decision_Tree_Classification = TRAIN_DECISION_TREE(CNN_Features)
Indexed_Content = Extract_Features_and_Metadata(Dataset)
Registered_Works = Load copyright-registered works
Content_Analysis = PERFORM_CONTENT_ANALYSIS(Indexed_Content, Registered_Works)
Risk_Assessment = TRAIN_RISK_ASSESSMENT_MODEL(Content_Analysis)
for each work in Indexed_Content do
    Risk_Category = Risk_Assessment.Classify(work)
    Output the Risk_Category for each work
end for
End
```

Table 4.1: Performance measure of proposed model

| Metric | Proposed Value (%) | Watermarking Algorithm value (%) |
|---|---|---|
| Overall Accuracy | 96 | 95 |
| Precision (Person - PER) | 90 | NA |
| Precision (Location - LOC) | 91 | NA |
| Precision (Organization - ORG) | 91 | NA |
| Recall (Facility - FAC) | 89 | NA |
| Recall (Geo-political Entity - GPE) | 93.02 | NA |
| Recall (Vehicle - VEH) | 85 | NA |

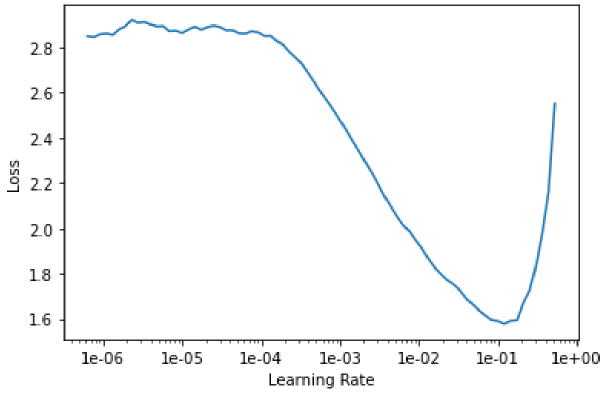

Fig. 4.1: Performance of Learning Rate at 0.01

is utilized to train the algorithm. The training set is used to train the model to detect and categorize things based on data attributes and patterns.

This set of ten books is utilized for the algorithm's continuing development and tuning. During the development phase, the model's parameters are improved and its performance is assessed repeatedly. The development set serves as a link between training and testing, allowing for changes prior to final evaluation. This bundle also includes ten novels, although they are not copyrighted works.

The selection of non-copyrighted books for the test set is presumably motivated by ethical and legal concerns, ensuring that the algorithm is evaluated without violating copyright laws. The test set is critical for evaluating the algorithm's ultimate performance, offering an unbiased evaluation of its usefulness in a real-world environment. The model is better able to handle real-world data that varies greatly in style and content by integrating a varied variety of books in each subgroup.

**4.2. Performance Measures.** The common metrics used for evaluating classification models are accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) is tested.

*Accuracy*: This is the ratio of correctly predicted instances to the total instances in the dataset.

*Precision*: This measures the ratio of correctly predicted positive observations to the total predicted positive observations.

*Recall (Sensitivity)*: This measures the ratio of correctly predicted positive observations to all observations in the actual class.

*F1-Score*: This is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

*ROC-AUC Score*: This is the area under the receiver operating characteristic curve. It is used to measure the model's performance across all classification thresholds. The performance is show in table 4.1 below.

Table 4.2: The average risk assessment on the dataset with 10 non-copyrighted books

| **Risk Assessment** | **Precision (%)** | **Recall (%)** | **F1-Score (%)** |
|---|---|---|---|
| High Risk | 92 | 88 | 90 |
| Medium Risk | 85 | 83 | 84 |
| Low Risk | 95 | 97 | 96 |

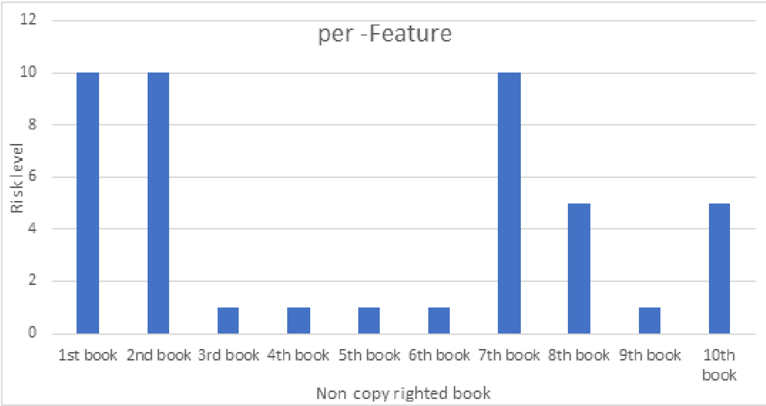


Fig. 4.2: Per Feature Risk Assessment Sample

**4.3. Risk Assessment Categorization.** Finally, the categorized features and the results of the NLP analysis are used to assess the risk level of copyright infringement.

1. *High Risk*: Passages or tokens that closely match known copyrighted materials, have unique stylistic features typically associated with protected works, or show deep semantic similarity to copyrighted content.
2. *Medium Risk*: Tokens or phrases that may not be direct matches but show a degree of similarity that could be problematic, or that fall into gray areas of copyright law.
3. *Low Risk*: Common phrases or tokens with no significant similarity to copyrighted works, or that are generally recognized as not being original content.

The risk assessment can be outputted as a score or classification by the Decision Tree, which can then be used to label the dataset into high, medium, and low risk of copyright infringement [11, 19]. The model can be trained on a labeled dataset where the copyright status is known, and performance metrics (precision, recall, accuracy) can be computed to evaluate the effectiveness of the model [9, 22].

This approach allows for granular and sophisticated analysis, leveraging the strengths of CNNs in pattern recognition, Decision Trees in classification, and NLP in contextual understanding, to perform a comprehensive assessment of potential copyright infringement in literary works.

The per feature-based risk analysis shown in below graph. The 10 non copy right book is assessed using per feature in the dataset. The 10-book person name is tested non copyrighted test set. Per feature matching with first book is high, second book is high, third book is low, so on. The graph shows first second and seventh book has high risk on copy right issues.

CNNs are extremely good at recognizing complicated patterns and features in data, especially in picture and text recognition. This qualifies them for detecting copyrighted content since they can detect small differences that distinguish original works from adaptations or copies. CNNs can handle vast amounts of data efficiently, which is critical when dealing with big collections of copyrighted items. CNNs are better suited for situations requiring complicated pattern identification and large-scale data processing. Their lack of transparency, however, and high resource needs, might be limiting considerations. Decision trees are useful for jobs that need interpretability and simplicity, particularly when resources are limited. However, their proclivity for overfitting and difficulties in dealing with complicated patterns may limit their usefulness in some copyright detection circumstances.

**5. Conclusion.** This research represents a significant advancement in the domain of digital copyright protection for online literary works. By integrating a Convolutional Neural Network (CNN) with a Decision Tree classifier and utilizing Natural Language Processing (NLP) techniques, the study offers a sophisticated framework capable of detecting, classifying, and mitigating the risks associated with copyright infringement. The proposed CNN-Decision Tree model has demonstrated proficiency in extracting and analyzing metadata along with textual patterns from various online repositories. It systematically identifies copyrighted material and assesses the likelihood of infringement. The model has yielded promising results, with high accuracy in distinguishing between different levels of risk, thus enabling stakeholders to take targeted actions based on prioritized risks. Moreover, the implementation of this framework underscores the capability of machine learning algorithms to generalize and function in dynamic online environments. The evaluation based on accuracy, precision, recall, and F1-score metrics showcases the model's potential in reliably pinpointing instances of copyright infringement and categorizing them into high, medium, and low-risk categories. As AI continues to intersect with copyright law, further research into the legal and ethical implications of automated copyright enforcement is necessary to ensure fair and just applications of the technology. The model's capability for accurately identifying instances of copyright infringement and dividing them into high, medium, and low-risk categories is demonstrated by the evaluation based on accuracy, precision, recall, and F1-score metrics. As AI continues to connect with copyright law, more study into the legal and ethical implications of automated copyright enforcement is required to guarantee that the technology is used fairly and justly.

## REFERENCES


[1] C. Anfray, B. Arnold, M. Martin, S. Eremenco, D. L. Patrick, K. Conway, C. Acquadro, I. Translation, and C. S. I. G. (TCA-SIG), *Reflection paper on copyright, patient-reported outcome instruments and their translations*, Health and Quality of Life Outcomes, 16 (2018), pp. 1–6.

[2] B. Bodó, D. Gervais, and J. P. Quintais, *Blockchain and smart contracts: the missing link in copyright licensing?*, International Journal of Law and Information Technology, 26 (2018), pp. 311–336.

[3] G. Carugno, *How to protect traditional folk music? some reflections upon traditional knowledge and copyright law*, International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 31 (2018), pp. 261–274.

[4] L. Che, *Copyright protection of literary works based on data mining algorithms*, Scientific Programming, 2022 (2022), pp. 1–10.

[5] P. Devarapalli, *Machine learning to machine owning: redefining the copyright ownership from the perspective of australian, us, uk and eu law*, Pratap Devarapalli,(2018). Machine Learning to Machine Owning: Redefining the Copyright Ownership from the perspective of Australian, US, UK and EU law, European Intellectual Property Review, 40 (2018), pp. 722–728.

[6] N. K. S. Dharmawan, *Protecting traditional balinese weaving trough copyright law: is it appropriate?*, Diponegoro Law Review, 2 (2017), pp. 57–84.

[7] H. B. Essel, R. B. LAMPTEY, and K. O. Asiamah, *Awareness of law students of kwame nkrumah university of science and technology (knust) on copyright law: Emphasis on photocopying and fair use.*, All Nations University Journal of Applied Thought, 6 (2019), pp. 71–87.

[8] M. Finck and V. Moscon, *Copyright law on blockchains: between new forms of rights administration and digital rights management 2.0*, IIC-International Review of Intellectual Property and Competition Law, 50 (2019), pp. 77–108.

[9] S. Geiregat, *Digital exhaustion of copyright after cjeu judgment in ranks and vasiļevičs*, Computer Law & Security Review, 33 (2017), pp. 521–540.

[10] T. He, *The sentimental fools and the fictitious authors: rethinking the copyright issues of ai-generated contents in china*, Asia Pacific Law Review, 27 (2019), pp. 218–238.

[11] E. Hudson, *The pastiche exception in copyright law: a case of mashed-up drafting?*, Intellectual Property Quarterly, 2017 (2017), pp. 346–368.

[12] V. Lunyachek and N. Ruban, *Managing intellectual property rights protection in the system of comprehensive seconday education*, Public Policy and Administration, 17 (2018), pp. 114–125.

[13] T. Margoni and M. Kretschmer, *Ai, machine learning and eu copyright law: A socio-legal analysis of ownership issues in training data*, in EPIP 2022, Date: 2022/09/14-2022/09/16, Location: Cambridge UK, 2022.

[14] R. Matulionyte, *Empowering authors via fairer copyright contract law*, University of New South Wales Law Journal, The, 42 (2019), pp. 681–718.

[15] J. P. McSherry, *The labor of literature: Democracy and literary culture in modern chile*, 2018.

[16] C. S. Myers, *Plagiarism and copyright: best practices for classroom education*, College & Undergraduate Libraries, 25 (2018), pp. 91–99.

[17] K. H. Nekit, H. O. Ulianova, D. O. Kolodin, and D. Kolodin, *Website as an object of legal protection by ukrainian legislation*, (2019).
[18] J. H. Rooksby and C. S. Hayter, *Copyrights in higher education: motivating a research agenda*, The Journal of Technology Transfer, 44 (2019), pp. 250–263.
[19] M. Sag, *The new legal landscape for text mining and machine learning*, J. Copyright Soc'y USA, 66 (2018), p. 291.
[20] S. Schroff, *An alternative universe? authors as copyright owners-the case of the japanese manga industry*, Creative Industries Journal, 12 (2019), pp. 125–150.
[21] N. H. Sharfina, H. Paserangi, F. P. Rasyid, and M. I. N. Fuady, *Copyright issues on the prank video on the youtube*, in International Conference on Environmental and Energy Policy (ICEEP 2021), Atlantis Press, 2021, pp. 90–97.
[22] W. Slauter, *Introduction: copying and copyright, publishing practice and the law*, Victorian Periodicals Review, 51 (2018), pp. 583–596.
[23] E. J. Tao, *A picture's worth: The future of copyright protection of user-generated images on social media*, Indiana Journal of Global Legal Studies, 24 (2017), pp. 617–636.
[24] U. F. Ugwu, *Reconciling the right to learn with copyright protection in the digital age: Limitations of contemporary copyright treaties*, Law and Development Review, 12 (2019), pp. 41–77.

*Edited by:* Rajanikanth Aluvalu
*Special issue on:* Evolutionary Computing for AI-Driven Security and Privacy: Advancing the state-of-the-art applications
*Received:* Dec 7, 2023
*Accepted:* Jan 3, 2024