Graph-Based Approach for European Law Classification

Raffaele Russo*, Giuliano Di Giuseppe*, Alessandro Vanacore*,
Valerio La Gatta+, Antonino Ferraro+, Antonio Galli+, Marco Postiglione+, Vincenzo Moscato+
Department of Electrical Engineering and Information Technology
University of Naples Federico II
Naples, Italy
*{raffaele.russo45, giu.digiuseppe, alessa.vanacore}@studenti.unina.it, +{name.surname}@unina.it

2023 IEEE International Conference on Big Data (BigData) | 979-8-3503-2445-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/BigData59044.2023.10386684

***Abstract*—Deep learning, owing to its transformative influence across a myriad of sectors, has recently made its foray into the legal domain, instigated by the surge in digitization. Among the multitude of applications in this space, legal document classification emerges as a pivotal yet complex undertaking. Legal texts, characterized by unique domain-centric semantics and intricate linguistic patterns, necessitate precision-driven classification systems for numerous practical implications. This paper illuminates the challenges and opportunities in automating the classification of European Union (EU) legal documents, emphasizing the interrelationships among statutes and the hierarchical nature of legal references. In this context, we introduce a novel graph data modeling technique that adeptly marries content-centric indicators with the relational dynamics inherent among diverse legal documents. Central to our approach is a framework that melds text embeddings with graph neural networks for the classification of legal documents aligned with their subject-based directories. Empirical evaluations on the EU law dataset underline the efficacy of our model across varying granularities, from general thematic categories to intricate subtopics. This endeavor not only augments the comprehensibility and accessibility of EU jurisprudence but also holds significant implications across regulatory compliance, legal research, and policy formulation, underscoring the potential of deep learning in reshaping legal paradigms.**

***Index Terms*—legal document classification, natural language processing, graph neural networks**

I. Introduction

Deep learning has profoundly impacted a myriad of sectors, driving unprecedented advancements in a variety of applications related to healthcare [1], predictive maintenance [2] and information authenticity [3]. This metamorphic leap is largely a consequence of the burgeoning data availability. In the legal domain, the evolution towards digitization has unlocked numerous prospects for automation, including tasks like law summarization [4], judgment prediction [5], and legal document classification [6], [7]. Yet, the journey of integrating automation in the legal realm is not devoid of challenges. While the objective of this transformation is to simplify processes, enhance efficiency, and curtail costs, obstacles remain especially in the context of legal document classification. Among the primary concerns are the meticulous handling of legal vernacular [8], the imperativeness of accuracy to deter severe legal misunderstandings [9], and the ethical deployment of algorithms [10]. A particularly daunting task lies in the translation of complex legal documents into a format decipherable by machines [11], which is a multifaceted challenge within the Natural Language Processing (NLP) ambit.

At the heart of these challenges is the unique nature of legal texts. These documents resonate with recurring themes, entrenched legal foundations, and specific linguistic patterns all underscored by a unique domain-centric semantics [12]. The intricate nature and nuances present in these documents necessitate precise classification for a multitude of purposes: from facilitating rapid access to pertinent laws and supporting policymakers in discerning legal interconnectedness, to providing invaluable assistance to legal professionals during case research [13]. An effective and precise classification system also stands to augment the transparency and navigability of the legal framework for the general public.

This narrative elucidates the latest interest around the issue of automatically classifying European Union (EU) legal documents based on their *subject-based* directories [14]. Indeed, a refined and accurate classification mechanism can redefine the landscape of regulatory compliance, elevate the efficiency of legal research, and bolster the efficacy of legal services, amongst other advantages. Yet, conventional techniques [15], grounded in content-centric analyses, often exhibit shortcomings, as they overlook the intricate interrelationships prevalent among diverse legal documents. Indeed, laws recurrently allude to one another to furnish context [16], invoke authority [17], or delineate procedural nuances [18]. Notable examples of these intricate interrelationships among legal documents include statutory cross-references, where one statute explicitly refers to another, as well as hierarchical references that manifest through various spatial strata and temporal contingencies. Spatial strata can be understood as the categorization of legal documents based on their jurisdictional applicability, ranging from regional and national to international levels. Temporal contingencies, on the other hand, refer to situations where a particular statute may supersede specific segments of another, creating a time-dependent hierarchy in the legal documentation.

Given these premises, in this paper, we design a simple, yet effective, graph data modelling procedure that integrates both content-driven cues and relational dynamics spanning diverse laws. Then, we propose a framework for automatically classifying legal documents based on their *subject-based* directories. At its core, our model synergizes text embeddings with graph neural networks to execute the task.

Our experiments on the EU law dataset [19] show promising performance at performing the task at different granularity levels. Specifically, our findings underscore the proficiency of our system not merely in discerning the overarching directory or thematic underpinning of the law (e.g., *Customs Union and free movement of goods*), but also its adeptness in pinpointing sub-facets (e.g., *Application of the Common Customs Tariff* and *International customs cooperation*).

Overall, our work seeks to bolster the accessibility and interpretation of EU jurisprudence. The ramifications of this study traverse the legal sphere, proffering dividends in regulatory compliance, juridical inquiries, operational and fiscal optimization, scalability, enhanced legal provisions, policy assessment, and data veracity.

II. Related Works

A. EU Legal Documents

The European Union's "Directory of Legal Acts" serves as a sophisticated numerical taxonomy, designed to offer structured, subject-centric access to the plethora of EU legislative instruments. This directory delineates three paramount categories: (i) Legal acts; (ii) International agreements; and (iii) Preparatory documents. Intrinsically structured, the directory manifests 20 principal chapters, each encapsulating a discrete realm of EU undertakings. A comprehensive list of these chapters, accompanied by their assigned primary indices, is elucidated in Table I.

Diving deeper, each principal chapter is meticulously segmented into multiple sub-chapters to encapsulate granular legislative nuances. For illustrative clarity, "Chapter 2 - Customs Union and Free Movement of Goods" undergoes further stratification to elucidate distinct legislative facets, as expounded upon in Table II. Notably, the taxonomical division within subchapters is non-mutually exclusive; a legislative document, while primarily categorized under a particular sub-chapter, might concurrently be associated with secondary sub-chapters within the overarching directory. As a result, the same legal document is designated to a primary chapter and could be associated with a principal sub-chapter as well as several secondary sub-chapters.

Navigating this intricate framework underscores the indispensable need for an automated classification mechanism tailored to both chapters and their nuanced sub-chapters. Such systematic categorization is pivotal in streamlining, deciphering, and organizing the expansive corpus of EU legal instruments, thereby bolstering an array of legal, regulatory compliance, and academic pursuits.

TABLE I: Primary Chapters

Id	Chapter
01	General, financial and institutional matters
02	Customs Union and free movement of goods
03	Agriculture
04	Fisheries
05	Freedom of movement for workers and social policy
06	Right of establishment and freedom to provide services
07	Transport policy 1006
08	Competition policy 1993
09	Taxation 258
10	Economic and monetary policy and free movement of capital
11	External relations
12	Energy
13	Industrial policy and internal market
14	Regional policy and coordination of structural instruments
15	Environment, consumers and health protection
16	Science, information, education and culture
17	Law relating to undertakings
18	Common Foreign and Security Policy
19	Area of freedom, security and justice
20	People's Europe

TABLE II: Subchapters of Chapter 02 - Customs Union and Free Movement of Goods

Id	Subchapter
02.05	General
02.07	Statistics
02.10	General customs rules
02.20	Basic customs instruments
02.30	Application of the Common Customs Tariff
02.40	Specific customs rules
02.50	Mutual assistance
02.60	Proceedings and penalties
02.70	International customs cooperation

B. Legal Artificial Intelligence

Historically, text classification predominantly utilized conventional machine learning techniques [20], including Support Vector Machines (SVM) [21] and Naive Bayes [22]. The subsequent advent of deep learning introduced advanced sequence architectures like Recurrent Neural Networks [23] and Long Short-Term Memory (LSTM) units [24]. Notably, the emergence of contextual transformer-based models, such as BERT [25] and RoBERTa [26], marked a significant evolution and has eventually led to the most recent foundation models such GPT [27] or LLaMa2 [28]. Alongside these advancements, there was a pronounced shift towards refining document embeddings, integrating both domain-tailored [29] and graph-based methodologies [30]–[32].

Given its inherent semantic intricacies and unique structural paradigms, the legal sector became a fertile ground for the application of graph-based methodologies. Such techniques demonstrated proficiency in diverse legal applications encompassing charge predictions [33], legal counsel generation [34], and advanced document retrieval mechanisms [35]. Seminal works [36], [37] accentuated the criticality of understanding and mapping relationships amongst legal textual entities. Augmenting domain-specific knowledge emerged as a pivotal strategy, bolstering classification precision. This was evident in models such as Graph LSTM [38] and in research endeavors that integrated heterogeneous graphs for the precise identification of legal statutes [39]. A salient initiative emanated from the EUR-LEX portal, where a nuanced multi-label classification challenge entailed the annotation of documents using the European Vocabulary in tandem with an array of deep learning constructs [40].

While the merit of graph modelling and graph neural networks in the legal arena is indisputable, our endeavor delineates from the extant literature in notable ways. Primarily, our exploration delves into the potential of graph modelling in executing ultra-granular classification tasks, encompassing not just the primary directory prediction but also possible sub-chapter affiliations of a legal document. Furthermore, our research scenario is imbued with added complexity as a single document might span multiple sub-chapters, necessitating a multi-label classification approach as opposed to the conventional single-label paradigm.

III. Data & Problem Formulation

A. Data Collection

In our study, we employ a pre-existing multilingual corpus [19] that curates a robust assembly of 57k EU legal directives, annotated as per the approximately 7k labels delineated in the EUROVOC[1] lexicon. While the dataset's magnitude might not seem extensive, its distinctive merit lies in its diversity — encapsulating a wide spectrum of juridical themes, document categories, and linguistic nuances. Our analysis predominantly concentrates on a subset of 11k English-language documents, from which we extracted the following information:

- CELEX_ID: Distinctive identifier for the legislation.
- Text: The substantive content of the legislation.
- Citation_ID: An intrinsic reference identifier for the legislation.
- Citations: Identifiers corresponding to other legislative directives alluded to within the document.
- Directory code: The primary EU chapter classification for the legislation, as cataloged in Table I.

For illustrative clarity, Table III offers exemplars of decisions instituted by the EU Council pertaining to nominations for supplementary committee constituents. It is worth noting that the secondary example explicitly cites the first, via the reference tag *OJL31,7.2.2015,p.25*. This citation underscores the superseding nature of the latter over the former, stemming from the cessation of the previous member's tenure and the ensuing requirement to induct a successor. Such intricacies accentuate the imperative to incorporate auxiliary metadata in tandem with the primary textual content for a holistic understanding of the legislative framework.

[1] https://eur-lex.europa.eu/browse/eurovoc.html

B. Problem Formulation

To elucidate the structural hierarchy and interconnections of EU legal documents, we aim to perform systematic classifications based on the text and normative citations embedded within. We delineate our classification objective into three distinct tasks:

Task 1: Hierarchical Chapter Classification This task gravitates towards classifying legal manuscripts per the 20 foundational chapters enumerated in the EU's "Directory of Legal Acts". Given that each directive is aligned unequivocally with a solitary chapter, we are dealing with a single-label classification challenge. An analysis of label distribution, as manifested in Figure 1, indicates a skewed propensity, with most documents from the first, third, eleventh, and eighteenth chapters.

Task 2: Granular Sub-Chapter Classification Pivoting towards a more nuanced label delineation, this task intends to classify texts based on the sub-chapters indexed within the "Directory of Legal Acts". These sub-chapters are discerned through a hyphen-linked annotation system, wherein a tag "01-07" indicates affiliation to the primary chapter and the seventh sub-chapter sequentially. For instance, Table II shows the sub-chapters of second directory. It is crucial to note that the taxonomical demarcation for this task is contingent upon the foundational chapter designation from Task 1. Of the 20 preeminent chapters, we have discerned 87 unique chapter-sub-chapter dyads. Given the pronounced imbalance observed in Task 1, particularly towards Chapter 1, this disproportion is inherently mirrored in Task 2.

Task 3: Multi-label Sub-Chapter Classification Retaining the focal point on sub-chapter categorization, this task transitions into a complex multi-label classification paradigm. In this context, a manuscript is potentially ascribed to a primary sub-chapter while simultaneously resonating with multiple ancillary sub-chapters. This stratification embodies the quintessential and intricate cross-referencing nature intrinsic to EU legal manuscripts. Analogous to its predecessors, a conspicuous imbalance is discerned favouring specific document classifications.

Table IV shows aggregated statistics for each task. Notably, while the chapter classification leverages the entirety of the dataset, the subsequent tasks harness a subset of 9,941 manuscripts from the original pool of 11,298, attributable to the absence of sub-chapter affiliations in the residual documents.

IV. Methodology

Figure 3 depicts the three-stage architecture underpinning our analytical framework. In the initial phase, the graph

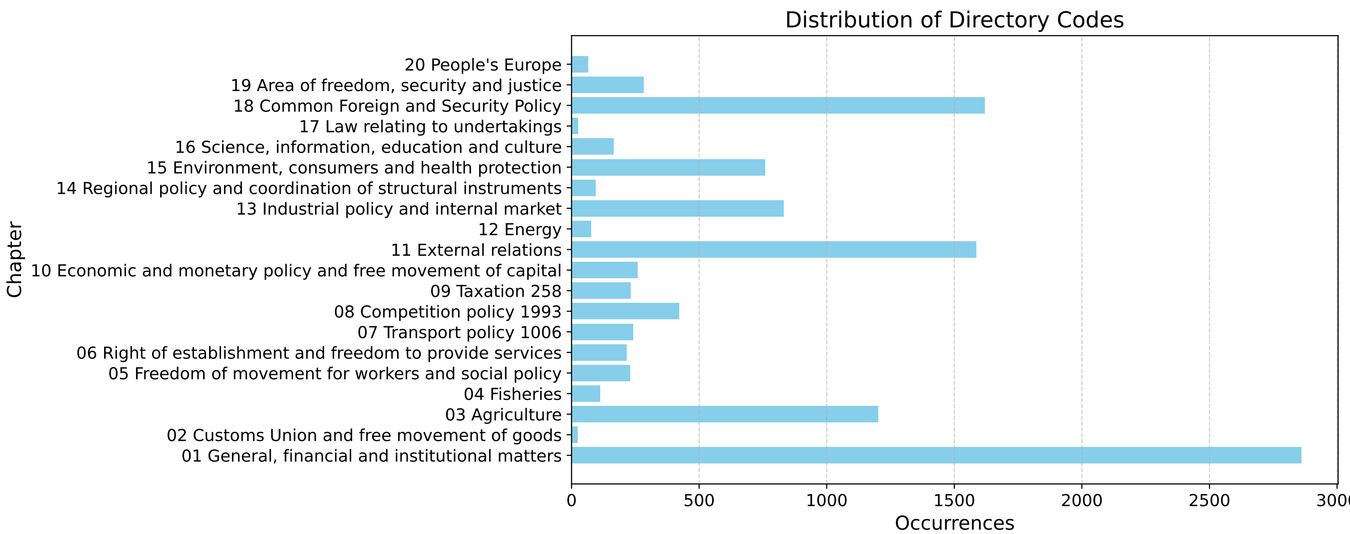

Fig. 1: Labels' Distribution for task 1

TABLE III: Some examples for the dataset. We highlight in bold the references that occurs within the depicted examples.

CELEX_ID	Citation_ID	Text	Citations
32015D0190	**OJL31,7.2.2015,p.25**	*COUNCIL DECISION (EU) of 5 February 2015 appointing two members and an alternate member of the Committee of the Regions* [...] *THE COUNCIL OF THE EUROPEAN UNION,* [...], *HAS ADOPTED THIS DECISION:* [...]	'OJL20,27.1.2015,p.42', 'OJL365,19.12.2014,p.143'
32019D0580	OJL100,11.4.2019,p.45	*COUNCIL DECISION (EU) of 8 April 2019* [...] *THE COUNCIL OF THE EUROPEAN UNION:* *Having regard to the Treaty on the Functioning of the European Union, and in particular Article 305 thereof,* *Whereas:* *(1) On 26 January 2015, 5 February 2015 and 23 June 2015, the Council adopted Decisions* **OJL31,7.2.2015,p.25** [...]	'OJL159,25.6.2015,p.70', '**OJL31,7.2.2015,p.25**', 'OJL20,27.1.2015,p.42'
32017D0614	OJL86,31.3.2017,p.13	*EU COUNCIL DECISION of 21 March 2017* [...] *The following is hereby appointed as a member of the Committee of the Regions* [...] *Mr Gerry WOOP, Staatssekretär für Europa (Land Berlin).* [...]	'OJL20,27.1.2015,p.42', 'OJL159,25.6.2015,p.70', '**OJL31,7.2.2015,p.25**'
32019D1107	OJL175,28.6.2019,p.37	*EU COUNCIL DECISION of 25 June 2019* [...] *appointing an alternate member, proposed by the Kingdom of Spain, of the Committee of the Regions* [...]	'OJL159,25.6.2015,p.70', '**OJL31,7.2.2015,p.25**', 'OJL332,18.12.2015,p.144', 'OJL20,27.1.2015,p.42'

TABLE IV: Dataset Overview

Classification	# istances	# classes	Typology
Chapter	11298	20	Single-label
Sub-chapter	9941	87	Single-label
Sub-chapter	9941	93	Multi-label

modelling component delineates a graphical representation encapsulating the entirety of the dataset under scrutiny. Subsequently, the data preprocessing module is entrusted with the extraction of contextual embeddings, representing the textual content inherent to each node. In the terminal phase, an advanced graph neural network performs the classification with node-level categorization.

We proceed to expound upon the individual constituents of this framework in the subsequent sections.

A. Graph Modelling

Our formulation is built upon a homogeneous graph $G(V, E)$, characterized by its nodes V and edges E as follows:

- **Nodes:** Every legal document constitutes a node. Specifically, a node encapsulates the textual content of a respec-

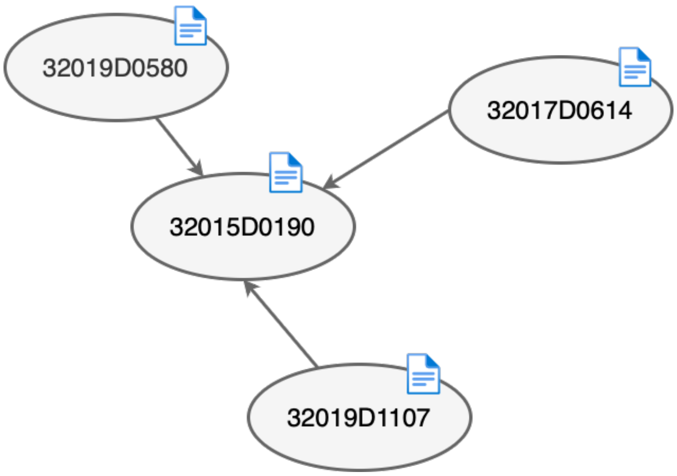

Fig. 2: Graph visualization of the examples showed in Table III. The identifiers are the CELEX_ID of the documents.

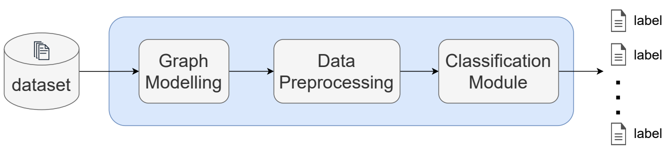

Fig. 3: The architecture of our framework

tive law. From a computation perspective, such content undergoes a transformation to yield vector embeddings, a process elucidated further in the subsequent data pre-preprocessing step. Predicated on the task at hand, a node may bear a singular label—indicative of the document's chapter or sub-chapter, or multiple labels, representing various sub-chapters to which the document is pertinent.

- **Edges:** Edges are symbolic of references or citations that a document extends to another. Formally, an edge $e = (v_i, v_j)$, with $v_i, v_j \in V$, signifies that the document encapsulated by v_i makes explicit reference to v_j within its textual framework.

Such a structural configuration succinctly embodies the intricate interplay amongst legal documents, mirroring the woven fabric of legal discourse. To illustrate, Figure 2 visualizes the interconnectedness of the four documents delineated in Table III. Through this graph, the symbiotic relationship between legal texts becomes manifest, underscoring the value of complementing textual analysis with relational insights.

Central to this modus operandi is the notion that proximate nodes within this graph likely resonate around a unified legal theme or subject. By capitalizing on these inherent topological affiliations, especially the citation linkages between documents, we can bolster the precision in discerning the legal context or category to which a node—and by extension, a legal document—belongs. Such a synergy of network structures enhances the fidelity in legal document classification by integrating relational cues into the decision-making paradigm.

B. Data Pre-processing

The objective of this module is the systematic transformation of textual data into a computationally amenable numerical format, priming it for subsequent analyses. Formally, given a document's text denoted as t, the objective is to derive a vector representation $vec_t \in \mathcal{R}^n$, where n represents the latent dimensionality. The process of text embedding is imperative for encapsulating the semantic intricacies inherent in legal prose, thereby enabling models to process and categorize content with heightened fidelity.

In practical terms, a multitude of embedding methodologies can be employed in this phase. The experimental section will subsequently elucidate the efficacy and implications of the chosen embedding paradigm.

C. Classification Module

Upon embedding the textual content, we transition to modeling the interrelations among entities based on the established graph schema. This network architecture is tailored to accommodate the inherent homogeneity and pronounced sparsity characteristic of the graph, thereby ensuring model efficacy and computational efficiency, even under rigorous scenarios. Subsequently, the classification module orchestrates the core task of categorizing the nodes.

Specifically, this module is underpinned by a sophisticated graph neural network, constituted by either one of the following pivotal layers:

- *GATConv (Graph Attention Network Convolution)* [41]: Integrating attention dynamics, this layer empowers nodes to differentially prioritize their adjacent nodes during information amalgamation. The essence of this selective prioritization rests on attention coefficients that are adaptively learned, thereby governing the weighting schema for information from adjacent nodes. Such a mechanism engenders a model with heightened adaptability and the finesse to discern intricate node interrelations.
- *GCNConv (Graph Convolutional Network Convolution)* [42]: This layer takes a more uniform approach, aggregating information from neighboring nodes with equal weighting. While this method might be less expressive compared to GATConv, it is generally more computationally efficient and simpler to implement.

The choice between GATConv and GCNConv hinges on the characteristics of the dataset at hand and the specific challenges posed by the legal document classification task. We will validate the effects of the graph layer in the experimental section.

In the final analytic phase, for each graph node, we derive a probabilistic vector indicating the likelihood of that node belonging to each of the possible classes. This is determined via a softmax activation, transmuting the raw class scores into a probabilistic distribution. Specifically, for the tasks encompassing chapter and sub-chapter classification, an argmax procedure is employed to identify the dominant class, thus ascribing the pertinent class label to the node. Conversely, for the multi-label sub-chapter classification task, the entire probabilistic output emanating from the softmax function is taken into consideration, facilitating an exploration of the model's nuanced predictive associations with the intrinsic labels of the document.

V. EXPERIMENTS

In this section, we describe the experimental protocol, the evaluation metrics and the results obtained for the tasks formalise in Section III-B.

A. Evaluation Metrics

In our evaluation methodology, we adopt standard metrics conventionally employed for classification tasks, encompassing precision, recall, F1-score, and accuracy. Nevertheless, given the inherent class imbalance present in our dataset, an over-reliance on accuracy as an evaluation metric may lead to models that are biased towards the majority class, potentially skewing the true performance of the classifier. Recognizing this limitation, we tailored our optimization strategy for the graph neural network to prioritize the macro-average F1-score across all classification tasks, as this metric provides a more holistic representation of the model's precision and recall capabilities.

B. Experimental Setup

Our dataset was partitioned utilizing a stratified sampling technique to ensure that each subset is representative of the entire dataset. Specifically, we allocated 60%, 20%, and 20% of the data to the training, validation, and test sets, respectively. To enhance the reliability and robustness of our evaluations, we employed Monte Carlo cross-validation, performing three independent training iterations and subsequently aggregating the outcomes to derive an average performance metric. This iterative approach is designed to mitigate the potential biases and variations inherent in a singular training instance.

1) Data Preproccesing: The preprocessing steps undertaken to prepare textual data for analysis encompassed:

- **Stop Word Removal:** Common words that typically do not convey substantial semantic importance were excluded to streamline the data and accentuate salient terms.
- **Lemmatization:** Words were transformed to their canonical form or lemma to engender consistency across the dataset, thus mitigating potential variations arising from inflections or derivations.
- **TF-IDF Vectorization:** Contrary to lemmatization, the Term Frequency-Inverse Document Frequency (TF-IDF) technique was employed to quantify the relevance of each term within a document relative to a corpus. This method facilitates feature extraction, offering a nuanced representation of text for subsequent analyses.

The resultant data, having undergone the aforementioned preprocessing stages, stands optimized for subsequent analytical tasks and model training. We will evaluate the effects of using the TF-IDF techinque to refine the representation rather than just relying on lemmatization.

2) Classification Module : Within the classification module, the graph neural network was initialized with stochastic weights, featuring a latent dimension of 1024. Optimization objectives varied per task: cross-entropy loss was employed for tasks 1 and 2, whereas the Jaccard loss, which measures the degree of overlap between predicted and actual label sets, was preferred for the multi-label nature of task 3. Training proceeded until model convergence, and preemptive measures (i.e., early stopping) were adopted to counteract potential overfitting, gauged by performance on the validation set. Specifically, we trained the model for 25 epochs with a batch size of 32. Finally, in the case of task 3, characterized by its multi-label nature, particular emphasis was placed on refining the threshold parameter. This parameter played a crucial role in establishing the criteria for validating a predicted class, shaping the precision of the model's multi-label predictions.

C. Experimental Protocol

Our experimental protocol is meticulously designed to ascertain the optimal parameter configuration tailored for each distinct task. This determination encompasses a comprehensive evaluation of key hyper-parameters, specifically:

- **Text Embedding Technique:** We examined two principal methodologies for this purpose. The first is a canonical representation achieved through *lemmatization*, and the second, a more sophisticated representation using the *TF-IDF* (Term Frequency-Inverse Document Frequency) vectorization approach.
- **Graph Convolution Layer:** Two potential graph convolutional strategies were assessed - the *GATConv* (Graph Attention Network Convolution) and *GCNConv* (Graph Convolutional Network Convolution) layers.
- **Network Depth:** This pertains to the depth of the graph neural network, signified by the cumulative number of layers it encompasses.

Through systematic variation and assessment of these hyperparameters, we aim to fine-tune our model, honing its performance for each specific task.

D. Results

Our comprehensive analysis, executed in alignment with the established experimental protocol, unearthed multifaceted insights regarding the influence of hyperparameters across three distinct classification tasks.

Table V shows the results for Task 1 and Task 2. The proposed model showcased promising performance for Task 1, registering an F1-score surpassing the 0.70 threshold. Conversely, the performance dynamics exhibited a shift for Task 2, which was marginally subdued. This attenuation in performance can be attributed to the heightened class imbalance and the increased number of categories inherent to Task 2. Notably, while the accuracy metric for both tasks appeared ostensibly analogous, it is a somewhat deceptive reflection, given the underlying class imbalances. This underscores the importance of F1-score as a more robust metric in such scenarios.

TABLE V: Performance on Task 1 and 2 with different configurations of hyper-Parameters (bold indicates the best results, underline the first runner up).

Task	Graph Layer	Text Encoder	Nro. Layers	Accuracy	F1	Precision	Recall
Task 1	GCN	Lemmatizer	1	0.857	0.725	0.739	0.710
		TF-IDF	1	0.873	0.693	0.683	0.670
		TF-IDF	2	0.770	0.640	0.546	0.480
	GAT	Lemmatizer	1	0.870	**0.750**	**0.768**	**0.790**
		TF-IDF	1	**0.881**	0.711	0.713	0.720
		TF-IDF	2	0.767	0.611	0.508	0.770
Task 2	GCN	Lemmatizer	1	0.7890	0.3917	0.3859	0.4332
		TF-IDF	1	0.8621	**0.4340**	**0.4596**	0.4452
		TF-IDF	2	0.7942	0.3661	0.3214	0.5888
	GAT	Lemmatizer	1	0.7930	0.3881	0.3789	0.4325
		TF-IDF	1	**0.8639**	0.4218	0.4349	0.4364
		TF-IDF	2	0.8088	0.3598	0.3149	**0.5919**

TABLE VI: Performance on Task 3 with Different Configurations of Hyper-Parameters (bold indicates the best results, underline the first runner up).

Graph Layer	Text Encoder	Nro. Layers	Accuracy (Training)	Accuracy (Validation)
GCN	Lemmatizer	1	0.9240	0.7835
	TF-IDF	1	**0.9257**	**0.7898**
	TF-IDF	2	0.8353	0.7315
GAT	Lemmatizer	1	0.8383	0.7328
	TF-IDF	1	0.9140	0.7820
	TF-IDF	2	0.8383	0.7308

Interestingly, increasing the depth of the network by increasing the number of layers consistently failed to bolster performance. In fact, the model predominantly reached its zenith with just a single hidden layer. This phenomenon is likely rooted in the *over-smoothing* problem [43], where the inherent message-passing mechanism of the graph layers inadvertently diminishes the prominence of the target node in relation to its neighboring nodes.

In the realm of text embedding, a dichotomous trend emerged. Lemmatization proved pivotal for optimizing performance in Task 1. Yet, when the lens shifted to Task 2, the TF-IDF features emerged as the frontrunners, underscoring their importance for this specific task.

Lastly, when examining the efficacy of the graph layers, a nuanced pattern was observed. The GAT layer, in comparison to the GCN layer, exhibited superior performance for Task 1. However, this differential advantage dissipated in the context of Task 2, where the GAT's performance effectively converged with that of the GCN layer. This results is somewhat counter-intuitive as Task 2 is more complex than task 1 but its better results are achieved with the simpler GNN layer.

Shifting our focus to Table VI, the performance metrics for Task 3 are showed. Analogous patterns to Task 2 were discerned, albeit with an additional layer of granularity provided by delineating results for both training and validation sets. This detailed exposition was pivotal to underscore a heightened susceptibility to overfitting in the context of Task 3.

Collectively, these results not only highlight the intricate interplay between hyperparameters and their task-specific efficacies but also emphasize the nuanced contingencies that determine the optimal configuration for each distinct task.

VI. Conclusions & Future Works

In this paper, we introduce a novel task centered around the automated classification of EU legal documents according to their *subject-based* chapters or sub-chapters. Our proposed framework melds the strengths of text embeddings with graph neural networks (GNNs) to adeptly navigate this classification landscape. Our graph data modeling strategy is meticulously designed, interweaving content-driven signals with relational dynamics characteristic of diverse legal documents.

In our comprehensive exploration of graph neural networks (GNNs) configurations across diverse tasks, we meticulously assessed variations in GNN layers and types, while concurrently investigating the impact of text encoding techniques. Our observations revealed that both the text encoding techqnique and GNN architecture can have important impact on the performance. However, there is no clear winner but the best performer depends on the specific data under analysis as well as the complexity of the task. On the contrary, we consistently observed that that deeper GNN models might introduce unnecessary complexity without enhancing performance but rather lead to the over-smoothing problem.

As we project future research trajectories, a compelling area of inquiry centers on the deep exploration of attention mechanisms embedded within GNNs as well as on the explainability of such techniques [44]. Broadening and potentially integrating diverse and interpretable text encoding techniques could potentially enhance model precision and recall. Furthermore, probing the adaptability, efficiency, and robustness of our proposed GNN configurations on expansive datasets remains crucial, offering insights into their practical applicability in real-world legal contexts.

ACKNOWLEDGMENT

This work has been partially funded by the Italian ministry of economic development, via the ICARUS (Intelligent Contract Automation for Rethinking User Services) project (CUP: B69J23000270005).

REFERENCES

[1] V. Moscato, G. Napolano, M. Postiglione, and G. Sperlì, "Multi-task learning for few-shot biomedical relation extraction," *Artificial Intelligence Review*, pp. 1–21, 2023.

[2] A. De Santo, A. Ferraro, A. Galli, V. Moscato, and G. Sperlì, "Evaluating time series encoding techniques for predictive maintenance," *Expert Systems with Applications*, vol. 210, p. 118435, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422015342

[3] T. Chakraborty, V. La Gatta, V. Moscato, and G. Sperlì, "Information retrieval algorithms and neural ranking models to detect previously fact-checked information," *Neurocomputing*, vol. 557, p. 126680, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231223008032

[4] Y. Kim *et al.*, "A legal document summarization using hierarchical attention networks," in *Proceedings of the International Conference on Artificial Intelligence and Law*, 2020.

[5] N. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos, "Predicting judicial decisions of the european court of human rights: a natural language processing perspective," *PeerJ Computer Science*, vol. 2, p. e93, 2016.

[6] I. Chalkidis and D. Kampas, "Deep learning in law: early adaptation and legal word embeddings trained on large corpora," in *International Conference on Artificial Intelligence and Law*. ACM, 2019, pp. 304–313.

[7] F. Amato, F. Colace, L. Greco, V. Moscato, and A. Picariello, "Semantic processing of multimedia data for e-government applications," *Journal of Visual Languages Computing*, vol. 32, pp. 35–41, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1045926X15000695

[8] D. M. Katz and M. J. I. Bommarito, "On the use and misuse of legal texts," *SSRN Electronic Journal*, 2017.

[9] H. Zhong, Z. Guo, C. Tu, Z. Liu, M. Sun, and J. Han, "Legal judgment prediction via topological learning," *arXiv preprint arXiv:1802.04085*, 2018.

[10] S. Wachter, B. Mittelstadt, and L. Floridi, "Why a right to explanation of automated decision-making does not exist in the general data protection regulation," *International Data Privacy Law*, vol. 7, no. 2, pp. 76–99, 2018.

[11] K. Branting *et al.*, "Transforming the law into a computational representation," in *Proceedings of the International Conference on Artificial Intelligence and Law*, 2020.

[12] K. D. Ashley, "Legal text analysis: The state of the art and the challenges ahead," *Artificial Intelligence and Law*, vol. 28, no. 2, pp. 123–132, 2020.

[13] N. Bansal, A. Sharma, and R. K. Singh, "A review on the application of deep learning in legal domain," in *Artificial Intelligence Applications and Innovations*, J. MacIntyre, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds. Cham: Springer International Publishing, 2019, pp. 374–381.

[14] J. Kerremans, R. De Busser, and B. De Clerck, "Automatic classification of documents within the european union's legal framework: A comparison between eurovoc descriptors and eu's subject matter directory codes," *International Journal on Artificial Intelligence Tools*, vol. 28, no. 02, p. 1950009, 2019.

[15] F. Bonchi and G. Casiraghi, "Challenges in content-based analysis and classification of legal documents," *Information Retrieval Journal*, 2019.

[16] M. Fragkoulis and M. Gagnon, "Extracting contextual information from legal texts for intelligent analysis and processing," *Artificial Intelligence and Law*, vol. 27, no. 4, pp. 461–489, 2019.

[17] V. Mak, J. Sanger, and F. Hogenboom, "Learning to invoke legal authority," in *International Conference on Artificial Intelligence and Law*. ACM, 2019, pp. 1–10.

[18] B. H. Smith, "Understanding legal texts: A cross-disciplinary approach to the mechanics of procedural interpretation," *Yale Law Journal*, vol. 130, no. 1, pp. 1–68, 2020.

[19] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, and I. Androutsopoulos, "Large-scale multi-label text classification on EU legislation," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 6314–6322. [Online]. Available: https://www.aclweb.org/anthology/P19-1636

[20] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[21] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[22] A. McCallum and K. Nigam, "A comparison of event models for naive bayes text classification," in *AAAI-98 workshop on learning for text categorization*, vol. 752. Madison, WI, 1998, pp. 41–48.

[23] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

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

[26] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 2019.

[27] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pretraining," 2018.

[28] J. Austin, X. Sun, R. Banerjee, M. Zaheer, G. Neubig, and I. Sutskever, "Li-lama2: Towards better language modeling and understanding," *arXiv preprint arXiv:2101.04473*, 2021.

[29] Y. Zhang, Q. Chen, Z. Yang, H. Lin, and Z. Lu, "Biowordvec, improving biomedical word embeddings with subword information and mesh," *Scientific data*, vol. 6, no. 1, pp. 1–9, 2019.

[30] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017.

[31] A. Kumar, P. Rai, and D. Halpenny, "Node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2020, pp. 855–864.

[32] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *6th International Conference on Learning Representations, ICLR 2018*, 2018.

[33] S. Chen, P. Wang, W. Fang, X. Deng, and F. Zhang, "Learning to predict charges for judgment with legal graph," in *Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series*, I. V. Tetko, V. Kůrková, P. Karpov, and F. Theis, Eds. Cham: Springer International Publishing, 2019, pp. 240–252.

[34] J. Yang, W. Ma, M. Zhang, X. Zhou, Y. Liu, and S. Ma, "Legalgnn: Legal information enhanced graph neural network for recommendation," *ACM Trans. Inf. Syst.*, vol. 40, no. 2, sep 2021. [Online]. Available: https://doi.org/10.1145/3469887

[35] J. S. Dhani, R. Bhatt, B. Ganesan, P. Sirohi, and V. Bhatnagar, "Similar cases recommendation using legal knowledge graphs," 2021.

[36] R. Ragesh, S. Sellamanickam, A. Iyer, R. Bairi, and V. Lingam, "Hetegcn: Heterogeneous graph convolutional networks for text classification," 03 2021, pp. 860–868.

[37] Q. Wang, K. Zhao, R. Amor, B. Liu, and R. Wang, "D2GCLF: Document-to-graph classifier for legal document classification," in *Findings of the Association for Computational Linguistics: NAACL 2022*. Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 2208–2221. [Online]. Available: https://aclanthology.org/2022.findings-naacl.170

[38] G. Li, Z. Wang, and Y. Ma, "Combining domain knowledge extraction with graph long short-term memory for learning classification of chinese legal documents," *IEEE Access*, vol. 7, pp. 139 616–139 627, 2019.

[39] S. Paul, P. Goyal, and S. Ghosh, "Lesicin: A heterogeneous graph-based approach for automatic legal statute identification from indian legal documents," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 10, pp. 11 139–11 146, Jun. 2022. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/21363

[40] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "Extreme multi-label legal text classification: A case study in eu legislation," 2019.
[41] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2018.
[42] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
[43] T. K. Rusch, M. M. Bronstein, and S. Mishra, "A survey on oversmoothing in graph neural networks," *arXiv preprint arXiv:2303.10993*, 2023.
[44] V. La Gatta, V. Moscato, M. Postiglione, and G. Sperlì, "Pastle: Pivot-aided space transformation for local explanations," *Pattern Recognition Letters*, vol. 149, pp. 67–74, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865521002014