Computer Systems Science and Engineering

Tech Science Press

DOI: 10.32604/csse.2024.043273 Published Online: 12 April 2024

ARTICLE

Multimodal Deep Neural Networks for Digitized Document Classification

Aigerim Baimakhanova[1,*], Ainur Zhumadillayeva[2], Bigul Mukhametzhanova[3], Natalya Glazyrina[2], Rozamgul Niyazova[2], Nurseit Zhunissov[1] and Aizhan Sambetbayeva[4]

[1]Department of Computer Engineering, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan

[2]Department of Computer and Software Engineering, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

[3]Department of Information and Computing Systems, Non-Profit Joint Stock Company Abylkas Saginov Karaganda Technical University, Karaganda, Kazakhstan

[4]Department of Information Systems, Al-Farabi Kazakh National University, Almaty, Kazakhstan

*Corresponding Author: Aigerim Baimakhanova. Email: aigerimbaimakhanova01@gmail.com

Received: 27 June 2023 Accepted: 14 November 2023

ABSTRACT

As digital technologies have advanced more rapidly, the number of paper documents recently converted into a digital format has exponentially increased. To respond to the urgent need to categorize the growing number of digitized documents, the classification of digitized documents in real time has been identified as the primary goal of our study. A paper classification is the first stage in automating document control and efficient knowledge discovery with no or little human involvement. Artificial intelligence methods such as Deep Learning are now combined with segmentation to study and interpret those traits, which were not conceivable ten years ago. Deep learning aids in comprehending input patterns so that object classes may be predicted. The segmentation process divides the input image into separate segments for a more thorough image study. This study proposes a deep learning-enabled framework for automated document classification, which can be implemented in higher education. To further this goal, a dataset was developed that includes seven categories: Diplomas, Personal documents, Journal of Accounting of higher education diplomas, Service letters, Orders, Production orders, and Student orders. Subsequently, a deep learning model based on Conv2D layers is proposed for the document classification process. In the final part of this research, the proposed model is evaluated and compared with other machine-learning techniques. The results demonstrate that the proposed deep learning model shows high results in document categorization overtaking the other machine learning models by reaching 94.84%, 94.79%, 94.62%, 94.43%, 94.07% in accuracy, precision, recall, F-score, and AUC-ROC, respectively. The achieved results prove that the proposed deep model is acceptable to use in practice as an assistant to an office worker.

KEYWORDS

Document categorization; deep learning; machine learning; classification; digitization

1 Introduction

Each day an increasing amount of massive data is generated on the Internet due to the regular development of social media, the Internet of Things (IoT), and mobile computing. Similar to

conventional data types, such a quantity of growing information demands a proper categorization to manage it most efficiently. Automatic document classification is becoming more crucial for trading companies, stores, and information retrieval as the volume of electronic text documents increases incredibly fast. Since it often satisfies the requirements of the present applications better, multi-label classification is noticeably preferable to single-label classification.

Vast amounts of generated digital data are available for data centers to use for quicker, more efficient, and automated processing in this digital age [1]. Property records examination, which generally deals with scanned images of distinct sorts, such as Certificates, Bank loans, Contracts, Payments, etc., is one of the sectors that provide ample opportunities for the application of Artificial Intelligence. The main challenge for automated administration and processing of the information hidden within these papers is classification. Properly handled property data directly influence the economic health of a nation. The fact that the materials are of diverse types, differing lengths, unstructured formats, and are closely connected in terms of the highlighted content [2] makes categorization challenging. The role of sophisticated exploratory data mining techniques, methodical exploration, and feature engineering has become crucial to develop a multi-class classifier and achieving an advanced level of accuracy in this assignment. Using the state-of-the-art embeddings methodology, which converts unprocessed text data into n-dimensional feature space, authors of some studies carefully examined property-related materials from six categories. They created a high-dimensional feature space using a linear kernel support vector machine (SVM) [3–5].

Modern approaches often include a variety of pre-processing techniques, such as feature selection, dimensionality reduction [6], and accurate document representations [7] to narrow the feature set while maintaining a high level of classification precision during the process. Nevertheless, this pre-processing technique has posed many severe problems, such as data loss, increased extra requirements analysis, task dependency or approval process, etc.

In artificial intelligence, neural nets with deep learning are now quite widespread, and it has been suggested that these networks surpass numerous state-of-the-art methodologies without the need for any parameter estimation. This is mainly fair when it comes to image analysis [8]. Still, it has also been proved that they perform better in Natural Language Processing (NLP), which includes tokenization, character segmentation, named entity identification, and semantic role labeling [9]. Nevertheless, whether the published work involves their use for categorizing multi-label documents has yet to be defined. Consequently, the application of neural networks to the problem of multi-label document categorization of university documents is the primary objective of this article. In the wake of the digital era, higher education institutions are grappling with an overwhelming amount of documentation that requires efficient and effective classification. Tackling this issue, the current study puts forward a deep learning-enabled framework for automated document classification. This proposal fills a pivotal niche in higher education, responding to the necessity for a structured and automated system to manage a burgeoning quantity of academic documentation.

The proposed method offers several distinct contributions to the field. Notably, it capitalizes on deep learning capabilities to enhance the document classification process, pushing the boundaries of existing methodologies. By taking advantage of state-of-the-art techniques, this study seeks to improve the accuracy and efficiency of document categorization.

The current work also addresses an understated aspect of automated classification: Interactivity. By fostering an interface that classifies and interacts intuitively with users, the proposed method augments user experience, a clear advantage over its predecessors. Evidence illustrating this enhancement will be provided in the ensuing sections.

Understanding the need for future adaptability and improvement, potential issues associated with the proposed method are also discussed in this paper. Future research directions are outlined to mitigate these issues, thereby ensuring the sustainability of this approach in the face of evolving requirements and technologies.

The following sections detail the construction and application of the proposed deep learning framework, its experimental validation, and how it compares to existing approaches. Furthermore, a comprehensive discussion will clarify its advantages, limitations, and potential areas for future enhancement, highlighting the study's significance and contributions to automated document classification in higher education. The novelty of the proposed research is the following:

A novel multimodal deep learning model has been developed in this research, which innovatively integrates visual and textual data for document classification. This advancement supersedes previous unimodal techniques, exploiting the benefits of both data types and achieving higher classification performance.

A unique feature extraction method has been introduced in the proposed model, maintaining the inherent structure and semantics of digitized documents. This methodology improves the accuracy of classification.

The challenge of handling imbalanced datasets, a frequent issue in document classification, has been addressed through a new method formulated in our research. This technique enhances the model's ability to classify under-represented classes, significantly boosting its versatility and applicability.

Furthermore, our study proposes an exhaustive evaluation framework for digitized document classification models. This approach considers the accuracy and interpretability of the model, promoting easier adoption in various practical applications.

The rest of this paper includes the following sections: Next section reviews the literature on categorizing digitized documents using deep learning methods. Section 3 deals with materials and methods, including the overall flowchart of the research, collected dataset, and proposed deep learning method for scanned document categorization. Following this, Section 4 provides obtained results, and at the end, the current research will be concluded, indicating the proposed method's advantages and research findings.

2 Literature Review

The categorization of documents is often accomplished via the use of artificial intelligence techniques. These techniques often employ labeled data to train classification models, and subsequently, the classifications are applied to texts that have not been labeled yet. The Vector Space Model, which typically depicts each document with a vector of all word occurrences weighted by their Term Frequency-Inverse Document Frequency (TF-IDF), is one of the widespread methods used in most published articles.

There have been several successful applications of classification techniques, such as Bayesian classifiers, Maximum Entropy (ME), Support Vector Machines (SVMs), and other similar approaches [10]. However, the most significant challenge presented by this job is the large dimensionality of the feature space inside the VSM, which reduces the accuracy of the classifier.

In recent times, "deep" Neural Networks (NN) have shown their better efficiency in a variety of NLP problems, particularly POS tagging, character segmentation, representation-based identification,

and semantic role labeling [11], all of which do not need any parametrization. Quite a few distinct architectures and learning methods were considered [12].

For instance, the researchers in [13] propose the utilization of two distinct convolutional neural networks (CNN) for tasks such as ontology construction, sentiment interpretation, and uni-label document categorization. Their networks comprise nine layers, with six being convolutional and the other three fully connected. Each layer has a distinct number of hidden units and a varied frame size. Applying both the English and the Chinese corpora, they demonstrate that the suggested strategy achieves much better results than the baseline methods (bag of words). Another fascinating piece of research [14] leverages pre-trained vectors from word2vec [15] in the first layer of their system. The authors determined that the suggested models perform better than the current state-of-the-art on four out of seven tasks, which include question categorization and sentiment analysis.

Traditional neural networks with numerous layers were also used for multi-label document categorization [16]. The authors have developed a unique error function and implemented it into a modified version of the classic backpropagation method for multi-label learning. The effectiveness of this method is examined using functional genomics and text classification. Within the subset of bottom-up methodologies, studies [17–19] highlight the efficacy of connected component analysis in tandem with the sliding-window technique. This synergy, coupled with the extraction of textural attributes, has exhibited notable success. These algorithms are relatively resistant to skew, although the processing they require may change depending on the selected metrics. According to reports, the accuracy of text detection ranges from 94% to 99%.

Problems in discriminating between many classes are often related to zone classifications. Haleem et al. [20] categorize eight representative items about the scanned texts. The experiment's findings have shown that successfully carrying out this activity is challenging. The reported mistake rate achieved 2.1%, yet 72.7% of the logos and 31.4% of the tables needed to be corrected. Although the other study showed a relatively high mean accuracy of 98.45%, 84.64% of trademarks and 72.73% of 'other' items were incorrectly identified [21].

An examination of the relevant published material reveals that most of the methodologies in question use contrast enhancement methods, only apply to specific types of documents, and combine hand-crafted features with multi-tiered methodologies [22]. The second observation revealed the need for more techniques to identify many fascinating elements that may be found in scanned texts. The main reason was that the properties of things have a great deal of variation regarding how they seem in different contexts.

Convolutional neural networks that can cope with these challenges using a particular approach were developed to overcome the abovementioned limits. This different study path is targeted at producing efficient classifiers rather than paying attention to constructing well-created low-level characteristics [23]. The challenge of translating model parameters, such as visual attributes, like the intensity of pixels, to specified outputs, such as an object type of object, is our task, and this study proposes a possible solution in this kind of situation. Due to the hierarchical structure in perceptual representation, the deep learning technique contains sets of inputs, hidden, and outputs. A model of this kind integrates high-level abstract notions with lower-level, more fundamental characteristics. One of the primary benefits of using such a strategy is the flexibility of learning. It is essential to emphasize that learning is not wholly innovative since it derives from traditional machine learning methods. Iterative data is supplied into the input of a deep neural network. Afterward, the training process allows the network to calculate layer by layer to create output, which is then compared with the right answer. Backpropagation is a process that corrects the weights of each node in a neural network

by moving the output error backward through the network. This helps to minimize overall error. This approach makes incremental improvements to the model while it is being computed. The convolution of a feature descriptor with input data is the most critical computational issue a CNN must solve. Therefore, the characteristics go from simple pixels to particular primitives such as lateral and vertical boundaries, circles, and color regions as they go up the hierarchy. On the other hand, CNN filters process all of the input vectors simultaneously, in contrast to traditional image filters, which only function on single-channel pictures. As a result of the multilayer filters' interpretation, they provide a robust response in regions in which a particular feature is encountered. Table 1 presents a comparative analysis of contemporary research in the field of document classification.

Table 1: Comparison of approaches for document classification problem

Reference	Problem	Applied method	Features	Dataset	Results
[24]	Automatic source scanner detection	1D CNN, Support vector machines	Handcrafted features	90 scanned documents from 9 scanners at	98.15% in training accuracy, 93.13% in test accuracy
[25]	Text and metadata extraction from scanned documents	Support vector machines and voting mechanism	Images features	–	–
[26]	Accreditation document classification	Naïve Bayes	Various images features	BCE-Arabic v1	94.4% accuracy
[27]	Automated detection of anomalies in imaged document scans	Traditional machine learning methods	–	–	0.900, 0.905, and 0.817 F1-score for different cases
[28]	Document classification		Autonomous features	Banknotes, crowd-sourced Mapping, VxHeaven dataset	99.85% accuracy

(Continued)

Table 1 (continued)

Reference	Problem	Applied method	Features	Dataset	Results
[29]	Digital document stream segmentation	Visual geometry group 16-Convolutional Neural Network (VGG16-CNN) and pre-trained bidirectional encoder representations from transformers (Legal-BERTbase)	Textual features	The real-time data set of 70,000 pages	97.37% F1-score and 97.15% accuracy
[30]	Classification of North and South handwritten scripts	KNN classifier	Gabor wavelet features	Dataset of 700 preprocessed images	92% accuracy

3 Materials and Methods

This section contemplates every aspect of our approach to subject identification in depth. Additionally, it is advised to undertake a textual and visual component-based study of the original document to develop and deploy a unique semantic topic identification technique. The rest of this section is devoted to a thorough description of the system architecture, focusing on the fundamental traits each module of the suggested framework share. In addition, it also provides extensive details regarding the multimedia knowledge base and the ontological model, which was employed as a basis for the base. Next, a comprehensive explanation of the presented Topic Detection method is provided.

Fig. 1 demonstrates the flowchart of the proposed system for the document categorization problem. A dataset comprising 8139 images, categorized into seven distinct classes, was collected for this study. Following this, DOM Parser was applied to facilitate accessing and modifying XML documents' style, structure, and content. Text documents were then preprocessed and added to the data repository. This process succeeded by collecting pertinent features required for training a categorization model. With the necessary features identified, a deep learning model was developed for training and testing. Consequently, categorized documents were distributed into distinct classes.

This allows the taxonomy classification to construct a classification taxonomy beginning with a concept. Comparisons have been drawn between the suggested measure, method, and baselines; the findings are illustrated and discussed in experimental Section 4 of the report.

Fig. 2 illustrates the proposed deep neural network to tackle the document categorization problem. The network's input constitutes a scanned document. This network comprises 11 layers, with the first layer functioning as a pooling layer. Subsequently, a Conv2D layer is utilized, and the resulting output is forwarded to the bottleneck layers. Within this network, there are five bottleneck layers. Following this, another pooling layer and a Conv2D layer are employed. The subsequent stage encompasses 128 neurons, while the penultimate layer comprises only seven layers, representing the document classes. The categorized class is ultimately determined using Maxpooling.

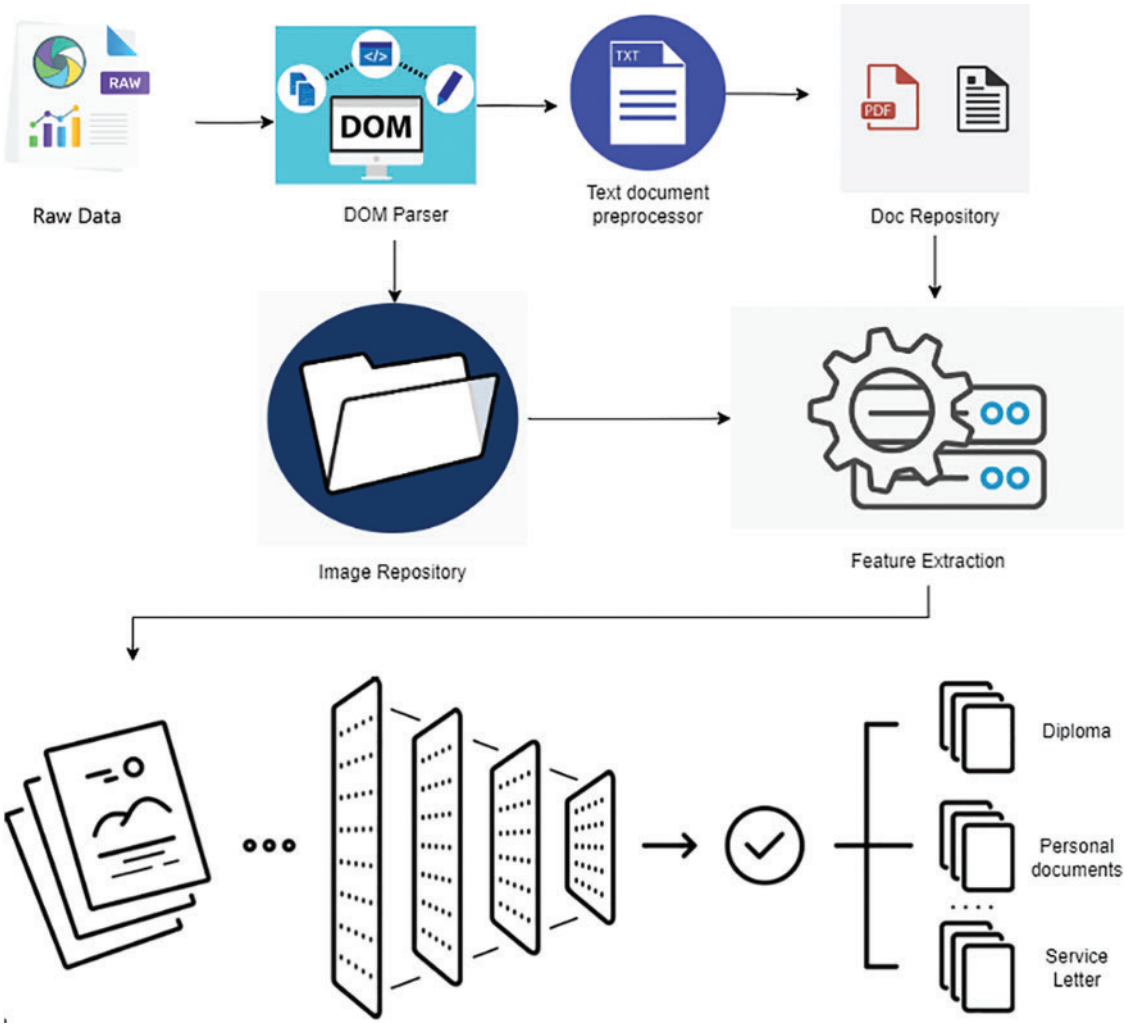

Figure 1: System architecture

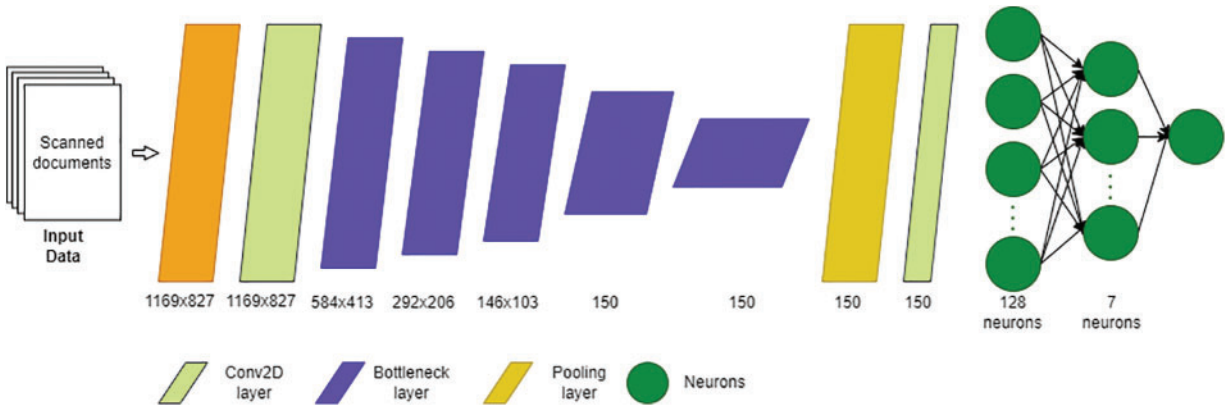

Figure 2: A multimodal classification system tailored for integrated text and image categorization. The training process seamlessly incorporates both linguistic and visual attributes

Fig. 3 offers a depiction of document samples drawn from the amassed dataset. Displayed in Fig. 3 are three distinct document types: A service letter, a diploma certificate, and a personnel document. This dataset comprises seven different types of university documents. With 8139 scanned documents spanning seven categories, the dataset occupies more than 4.7 Gb of storage.

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҮКІМЕТІ

ПРАВИТЕЛЬСТВО РЕСПУБЛИКИ КАЗАХСТАН

010000, Нур-Султан к., Үкімет үйі

010000, город Нур-Султан, Дом Правительства

№ исх: 21-09/875 дз от: 12.03.2021
№ вх.943//21-09/875.из/ДЗ-38 от: 12.03.2021

Депутатам
Мажилиса Парламента
Республики Казахстан,
членам фракции
«Народной партии Казахстана»
(по списку)

На № ДЗ-38 от 10 февраля 2021 года

Уважаемые депутаты!

Рассмотрев Ваш депутатский запрос относительно создания Межведомственной комиссии по совершенствованию мер социальной поддержки населения, пострадавшего вследствие ядерных испытаний на Семипалатинском полигоне, сообщаю следующее.

В настоящее время объем и характер оказания социальной поддержки граждан, пострадавших вследствие ядерных испытаний на Семипалатинском испытательном ядерном полигоне *(далее – СИЯП)*, зависят от зоны радиационного поражения и периода проживания на нем.

В целях совершенствования мер социальной защиты граждан, пострадавших вследствие ядерных испытаний на СИЯП, Министерством труда и социальной защиты населения Республики Казахстан проведен ряд исследований *(«Ретроспективный и проспективный анализ медицинских и радиоэкологических последствий бывшего Семипалатинского испытательного ядерного полигона для населения, проживающего к нему территорий», «Оценка эффективности адресных и реабилитационных мероприятий для населения, подвергшегося действию ионизирующего излучения в результате деятельности Семипалатинского испытательного ядерного полигона, и выработка предложений по их совершенствованию»)*.

По результатам проведенных исследований предложено оказывать медико-социальную помощь в зависимости от величины индивидуальной дозы облучения.

Данный подход позволит обеспечить оказание адресной социальной поддержки согласно полученным индивидуальным дозам облучения, а также с учетом доз облучения, полученных родителями.

Величина дозы облучения является модифицирующим фактором радиационного риска, которая зависит от особенностей человека, в том числе возраста на момент испытаний, продолжительности облучения. Следовательно, осуществлению каждого мероприятия по социальной защите должна

(a) Sample of a service letterdocument

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ

РЕСПУБЛИКА КАЗАХСТАН

АВТОРЛЫҚ ҚҰҚЫҚПЕН ҚОРҒАЛАТЫН ОБЪЕКТІЛЕРГЕ ҚҰҚЫҚТАРДЫҢ МЕМЛЕКЕТТІК ТІЗІЛІМГЕ МӘЛІМЕТТЕРДІ ЕНГІЗУ ТУРАЛЫ

КУӘЛІК

2018 жылғы «12» желтоқсан № 928

Авторлың (лардың) жеке , аты, әкесінің аты (егер ол жеке басын куәландыратын құжатта көрсетілсе):

ДАНИЯРОВ ТАЛГАТ АБУБАКИРОВИЧ, ЖОЛДАСБЕКОВА БИБИСАРА АБДИМАНАТОВНА, СЕМБИЕВ КУРМАНГАЗЫ ЗАКИРОВИЧ

Авторлық құқық объектісі: әрбір түрлері

«IN HIGHER EDUCATIONAL INSTITUTIONS CONTENT OF EDUCATION AND TRAINING CREATION

Объектінің атауы: OF TECHNOLOGY»

Объектіні жасаған күні: 03.09.2018

СВИДЕТЕЛЬСТВО

О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ

№ 928 от «12» декабря 2018 года

Фамилия, имя, отчество: (если оно указано в документе, удостоверяющем личность) автора (ов):

ДАНИЯРОВ ТАЛГАТ АБУБАКИРОВИЧ, ЖОЛДАСБЕКОВА БИБИСАРА АБДИМАНАТОВНА, СЕМБИЕВ КУРМАНГАЗЫ ЗАКИРОВИЧ

Вид объекта авторского права: литературное произведение

«IN HIGHER EDUCATIONAL INSTITUTIONS CONTENT OF EDUCATION AND TRAINING CREATION

Название объекта OF TECHNOLOGY»

Дата создания объекта 03.09.2018

Подписано ЭЦП Осанов Е. К.

(b) Sample of a diploma certificate

Figure 3: Examples of document categories within the curated dataset

4 Experiment Results

This section presents the results derived from the conducted experiments, offering a comprehensive overview of the performance and efficacy of the multimodal deep learning model. Detailed analyses have been conducted on diverse datasets, allowing us to evaluate the robustness and versatility of our methodology. Our findings and a thorough comparison with other existing techniques offer valuable insights into our proposed model's strengths and areas for improvement. The study mainly entails seven primary categories of university documents included in the cases handled by the STF. Below is a listing of these categories, with their original labels preserved.

4.1 Experimental Setup and Dataset

It should be noted that the legal cases include various papers, all of which were categorized under the heading "Miscellaneous." In this context, an annotation tool was developed and subsequently

employed by a group of four attorneys to manually categorize a collection of 8,139 documents. The percentage of papers that belong to each category is broken out graphically in Fig. 4, which may be seen below.

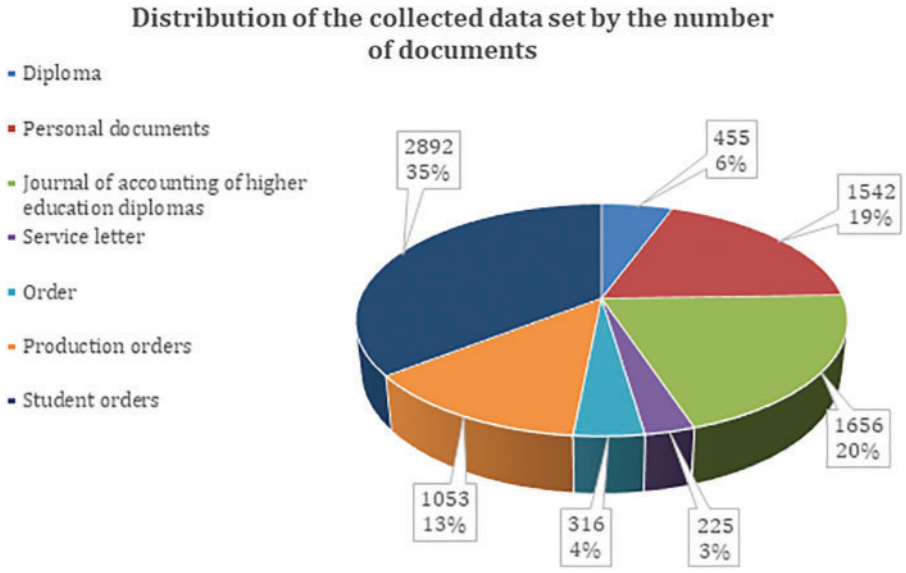

Figure 4: Categorization of document types within the dataset

It is standard practice to divide datasets into three sections to train and assess machine learning systems [31]. These divisions of the dataset are identified as the train, validation, and test subsets. Stratified splits are applied for each document class, thereby ensuring the inclusion of proportional class samples in each subset. The following ratios were employed, which are broken out further in Fig. 5:

- 70% is allocated for the training subset,
- 20% is designated for validation purposes, and
- 10% is reserved for the testing subset.

Fig. 4 presents the dispersion of the amassed dataset based on document quantity. The dataset comprises seven distinct categories of scanned materials, which are not balanced. Predominantly, the dataset is enriched by student-related directives, accounting for approximately 35% of the total scanned entities. Subsequently, the Journal of Accounting for advanced academic credentials, individual records, and manufacturing directives represent 20%, 19%, and 13% of the entire dataset, respectively. Notably, general orders, manufacturing directives, and correspondence related to services occupy the smaller segments of the dataset, contributing 6%, 4%, and 13%, respectively.

All the collected data composes of about 4.7 Gb of scanned pdf data. Fig. 5 illustrates the partition of these data by their volumes for each categorization. Diplomas, Personal documents, and higher education diplomas constitute 87.8% of all dataset volumes, with 42.6%, 19%, and 26%, respectively. The other four types of collected documents compose 12.2%.

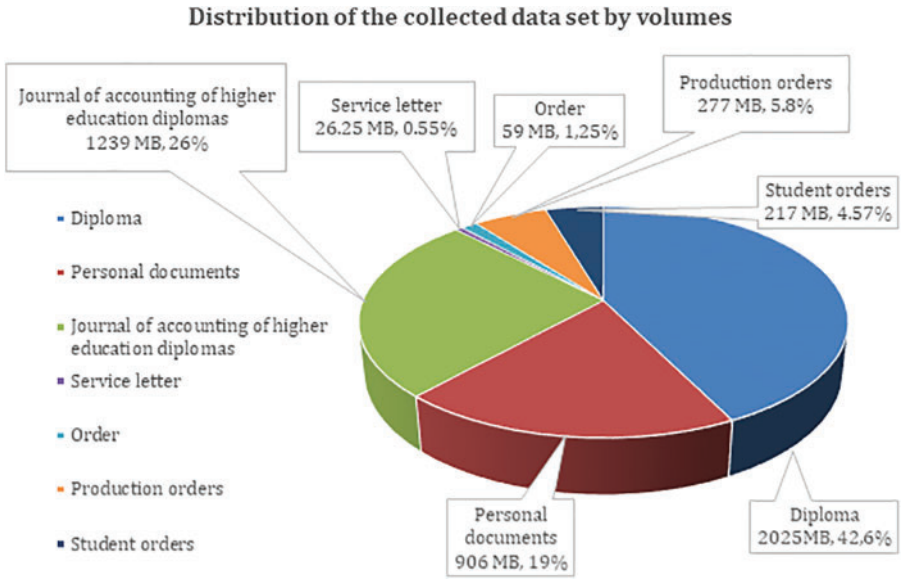

Figure 5: Distribution of the accumulated data set based on the occupied disk space

Fig. 6 showcases the allocation of the training, validation, and test subsets utilized for training the advanced deep learning model. The amassed dataset was divided into three parts: Training, validation, and test sets, comprising 70%, 20%, and 10% of the total data, respectively. Thus, dividing the dataset into three parts allows us to get high accuracy and understand whether the proposed model is effective in practice.

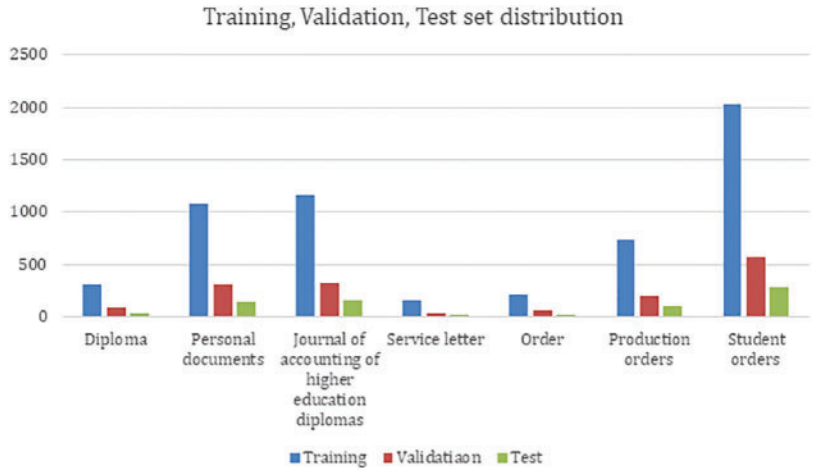

Figure 6: Distribution of training, validation, and testing subsets across each document category

Table 2 summarizes Figs. 5 and 6, illustrating the dispersion of the amassed dataset based on the count of scanned visuals and the data volume for every distinct category. It allows us to understand the relation of the number of images to the volume quality of the document by type. Thus, the model's training can commence following the dataset's preparation.

Table 2: Dataset overview

Type of the document	Quantity	Storage capacity
Diploma certificate	455	2025 MB
Personal documents	1542	906 MB
Academic ledger for advanced educational credentials	1656	1.21 GB
Correspondence related to services	225	26.25 MB
Order	316	59.5 MB
Production orders	1053	277 MB
Academic directives for students	2892	217 MB

4.2 Evaluation Parameters

This section aims to elucidate the evaluation parameters used to assess the proposed model and facilitate its comparison with other machine learning models. The following five indicators were selected as evaluation parameters: Accuracy, precision, recall, F-score, and area under the curve receiver operating characteristics (AUC-ROC). Eq. (1) stands for the formula of accuracy.

$$Accuracy = \frac{TP + TN}{P + N} \tag{1}$$

Eq. (2) shows the formula of precision.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Eq. (2) indicates the formula of recall.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Eq. (2) demonstrates the formula of the F-score.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

4.3 Results

This section presents the evaluation results of the proposed deep learning model for classification of scanned documents. Subsequently, the proposed model's model accuracy, validation accuracy, and confusion matrix are presented. Fig. 7 illustrates the accuracy of the model over the training of 100 epochs. The findings indicate that the suggested model exhibits commendable precision during both training and testing phases.

Fig. 8 depicts the model's loss across 100 epochs. The data suggests a swift decline in loss as the epochs increase. By the 80th epoch, both training and testing losses register below 0.3. This underscores the model's suitability for validated instances and its potential for effectively addressing automated document classification challenges.

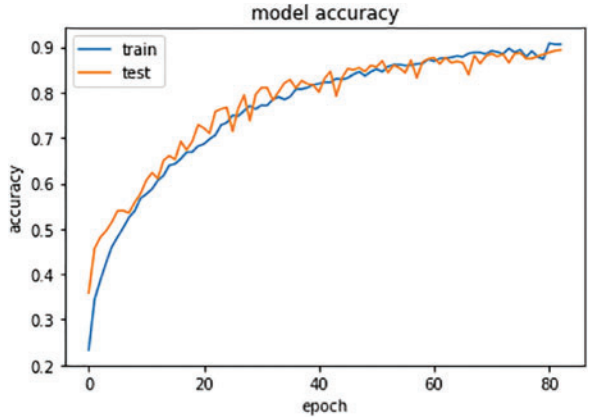

Figure 7: Accuracy of the model

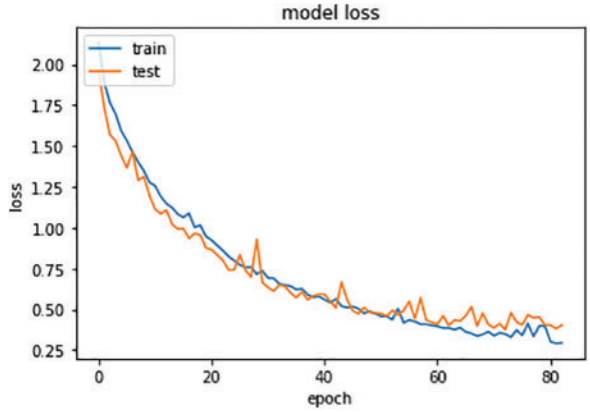

Figure 8: Model loss progression chart

Fig. 9 presents the confusion matrix corresponding to each categorized document type, serving as a performance assessment metric for the machine learning classification task that involves multi-class outputs. The table underneath contains the four possible permutations of anticipated and actual values. The results indicate that there is a minimum number of errors and confusion. In most cases, scanned documents are categorized correctly.

Table 3 contrasts the advanced deep learning model proposed with other established machine learning methodologies. The results claim that the proposed model gives the highest performance comparing the other methods in each evaluation parameter. The advanced model proposed yields performance metrics of 94.84% in accuracy, 94.79% in precision, 94.62% in recall, 94.43% in F-score, and 94.07% in AUC-ROC. These figures indicate the model's real-world applicability, and the provided dataset is apt for training both machine learning and deep learning frameworks in addressing document classification challenges.

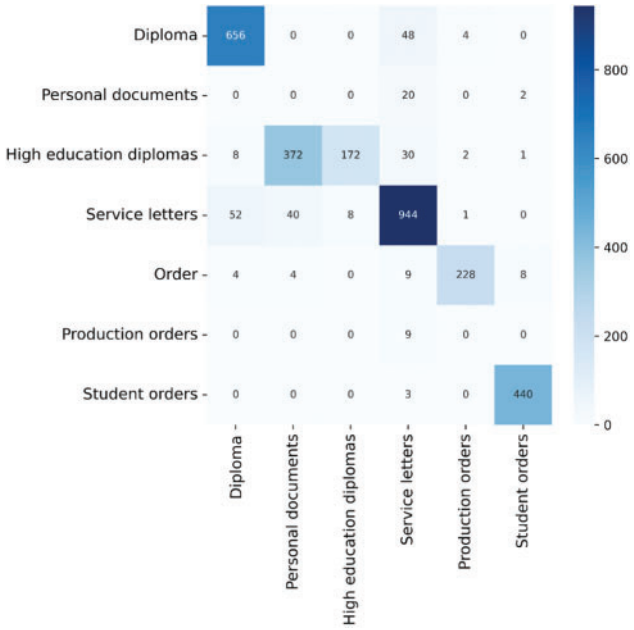

Figure 9: Confusion matrix

Table 3: Assessment of the model

Model	Accuracy	Precision	Recall	F-score	AUC-ROC
The proposed model	**94.84%**	**94.79%**	**94.62%**	**94.43%**	**94.07%**
Random Forest	82.73%	82.13%	82.34%	81.12%	81.09%
XGBoost	81.77%	81.31%	81.37%	81.21%	81.17%
Support vector machine	82.36%	82.17%	82.06%	82.21%	82.11%
Multilayer perceptron	80.67%	80.54%	80.51%	80.28%	80.12%
Decision trees	76.45%	76.37%	76.28%	76.17%	76.19%

Fig. 10 juxtaposes the advanced deep learning model proposed with five conventional machine learning methods, specifically: XGBoost, multilayer perceptron, random forest, support vector machines, and decision tree, all targeting document classification tasks. Multiple evaluation metrics, including accuracy, precision, recall, F-score, ROC-AUC, and threshold, are employed for this comparison. The results suggest that the advanced deep learning model consistently outperforms across all these metrics. As a result, this deep model proves to be effective for classifying academic institutional documents.

Table 4 compares the proposed method with state-of-the-art studies. Different studies have been developed for scanned document categorization of electronic health records, magnetic resonance images, and handwritten documents. The proposed Conv2D model categorizes educational documents of seven types with 94.84% accuracy.

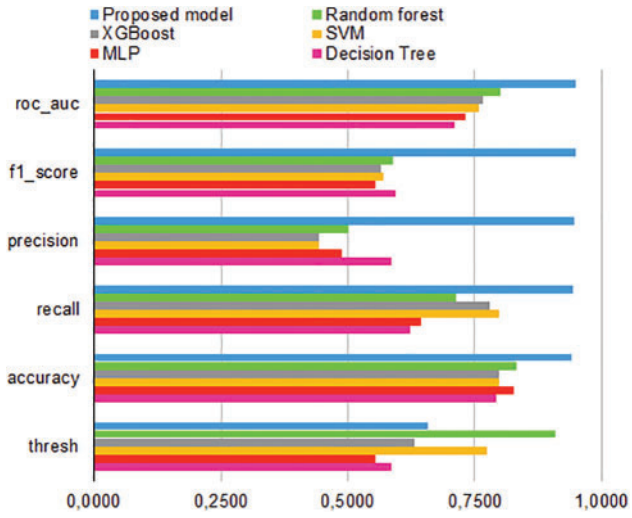

Figure 10: Derived outcomes juxtaposed with traditional machine learning techniques

Table 4: Evaluation of the acquired findings in contrast to the latest state-of-the-art research

Study	Approach	Document type	Dataset	Results
Current study	**Suggested Framework**	**8139 scanned documents of 7 types**	**Own dataset**	**Metrics recorded are 94.84% for accuracy, 94.79% in precision, 94.62% for recall, 94.43% concerning F-score, and 94.07% for AUC-ROC**
[32]	ClinicalBERT	A compilation of 2,988 scanned PDF visuals, originating from 955 distinct reports	Sleep analysis reports sourced from a prior research initiative at the University of Texas	AUROC of 0.95, document accuracy of 91.61%
[33]	An analysis approach devoid of learning, combined with a hybrid methodology, tailored for handwritten archival manuscripts	955 scanned sleep study reports	Two datasets (38 historical manuscripts and 51 historical manuscript pages)	Up to 98.5% segmentation rate
[34]	A fully automatic computational approach	250,000 historical letters	Xanthosine methyltransferase (XMT) datasets	82.06%

(Continued)

Table 4 (continued)

Study	Approach	Document type	Dataset	Results
[35]	Two-level transfer CNN model	Magnetic resonance imaging	Datasets from the Open Access Series of Imaging Studies coupled with resources from the Alzheimer's Disease Neuroimaging Initiative	92.30% accuracy
[36]	Automated software for 3D body scan measurements.	A collection of 3D scans encompassing 625 men aged 35–64, with 173 scans specifically categorized as 'abdominally obese.'	SizeKorea dataset	92% accuracy
[37]	Automated Paper Fingerprinting (APF) method	306 blank paper images	scanner image dataset from Universiti Kebangsaan Malaysia	97.07% accuracy
[38]	Deep Transfer Learning	Deep Transfer Learning		Accuracy: 0.8920, F1-score: 0.8751, Precision: 0.9281, Recall: 0.8279
[39]	Multi-Layered Perceptron	Multi-Page Digital Documents	The digital image document repository of the TI company has expanded to encompass around one billion documents	Recall value of 0.8927; Precision value of 0.9030; F1-score of 0.9380

In the pursuit of empirical validation, the proposed computational model was rigorously evaluated against the benchmark ICDAR 2017 competition dataset [40]. The comparative analysis is systematically presented in Table 5, where the performance metrics of the model are juxtaposed with those derived from the aforementioned dataset. This analytical juxtaposition corroborates the robustness of the proposed model, manifesting its competence in transcending dataset-specific constraints. The results unequivocally indicate that the model does not only retain its efficacy across diverse datasets but also achieves commendable evaluation scores in the domain of scanned document categorization, thus reinforcing the potential for wide applicability and the generalizability of its algorithmic constructs.

Table 5: Comparison of the proposed model with the other studies on ICDAR 2017 dataset

Study	Approach	Results
Current study	**Proposed deep neural network**	**91.9% accuracy, 91.7% precision, 97.5% recall, 91.7% AUC-ROC**
[41]	Cascade Mask R-CNN	91.8% precision, 91.6% recall, 91.7% F-score
[42]	Feature Pyramid Network	86% precision, 87.7% recall, 86.8% F-score
[43]	Faster RCNN	88.8% precision, 91.6% recall, 90.2% F-score
[44]	Cascade Network	93.5% precision, 33.1% recall, 48.9% F-score
[45]	LSTM Network	92% recall

5 Conclusion

In the current era, where the challenge is to manage an overwhelming amount of data, there exists a pressing need for AI technologies capable of classifying documents. During the extraction of features, such tools need to be able to swiftly and efficiently offer access to the searched data. To address this issue, our study proposes a method for identifying associated metadata. In contrast to other works, it incorporates verbal and visual elements into its structure. In addition, the study uses statistics and semantics to recognize the topics of multimedia online publications. In particular, using semantics enhances the value of our findings by elevating them to a higher level and immediately addressing a problem already well-known as the semantic gap. The proposed strategy yields high performance in various circumstances since it can assist in the organization of a document collection and describe the primary topic of a document through semantic multimedia analysis among concepts. Both of these capabilities contribute to the achievement of high-level performance. Consequently, an array of tests was executed to evaluate the effectiveness of various topic identification tasks, with the outcomes discussed in detail. In this regard, these results substantiate the superiority of the presented method for text topic detection compared to state-of-the-art methods concerning the subject at hand, such as LSA and LDA. Regarding visual topic identification, numerous types of descriptors have been rigorously tested. The most promising findings were selected, particularly those based on the characteristics derived from the activation layer of the deep neural network. Moreover, it was demonstrated that combining the strategy for textual topic identification with that of visual topic identification can feasibly enhance the overall task by leveraging the most successful elements of both strategies. Moreover, it can adjust the suggested architecture to implement multiple models for topic identification since the system is modular and can be reused. This task can be accomplished by adding new modules or updating the already present ones. The system has successfully passed all tests using a representative sample of online documents. However, the system's architecture permits the use of various libraries containing multimedia materials.

As part of our ongoing research, our next objective is to expand the evaluation of the proposed framework using a variety of document sets shortly. In addition, considering how difficult it is to locate data sets for websites with both text and multimedia elements, one of the following tasks will be constructing such a dataset while simultaneously using our knowledge base to identify discussion topics. Furthermore, several methods of subject identification will be investigated in the following days. Addressing the problem of multilingualism to enhance the performance of textual topic recognition is an intriguing area of study that needs to be pursued. In particular, the computational efficiency of the

textual topic recognition method will be the primary focus of our attention. To advance the current state of the art in this field, there are plans to research innovative approaches and algorithms for visual topic recognition.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization, A.B.; methodology, A.B., A.S., and B.M.; software, A.B.; data curation, A.B.; writing—original draft preparation, A.B., and A.S.; writing—review and editing, A.B., N.G., R.N., N.Z.; supervision, A.Z.

Availability of Data and Materials: The data used in this paper can be requested from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] H. Jain, S. Joshi, G. Gupta, and N. Khanna, "Passive classification of source printer using text-line-level geometric distortion signatures from scanned images of printed documents," *Multimed. Tools Appl.*, vol. 79, no. 11, pp. 7377–7400, 2020.

[2] D. Sultan *et al.*, "A review of machine learning techniques in cyberbullying detection," *Comput. Mater. Contin.*, vol. 74, no. 3, pp. 5625–5640, 2023.

[3] G. Wiedemann and G. Heyer, "Multi-modal page stream segmentation with convolutional neural networks," *Lang. Resour. Eval.*, vol. 55, no. 1, pp. 127–150, 2021.

[4] S. Chattopadhyay *et al.*, "MTRRE-Net: A deep learning model for detection of breast cancer from histopathological images," *Comput. Biol. Med.*, vol. 150, pp. 106155, 2022.

[5] A. Basu, K. Sheikh, E. Cuevas, and R. Sarkar, "COVID-19 detection from CT scans using a two-stage framework," *Expert Syst. Appl.*, vol. 193, pp. 1–14, 2022.

[6] B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, and M. Khassanova, "A skeleton-based approach for campus violence detection," *Comput. Mater. Contin.*, vol. 72, no. 1, pp. 315–331, 2022.

[7] R. das Neves, E. Lima, B. Bezerra, C. Zanchettin, and A. Toselli, "HU-PageScan: A fully convolutional neural network for document page crop," *IET Image Process.*, vol. 14, no. 15, pp. 3890–3898, 2020.

[8] E. Chaowicharat and N. Dejdumrong, "A step toward an automatic handwritten homework grading system for mathematics," *Inf. Technol. Control*, vol. 52, no. 1, pp. 169–184, 2020.

[9] I. Nasir *et al.*, "Pearson correlation-based feature selection for document classification using balanced training," *Sens.*, vol. 20, no. 23, pp. 1–18, 2020.

[10] W. Zaaboub, L. Tlig, M. Sayadi, and B. Solaiman, "Neural network-based system for automatic passport stamp classification," *Inf. Technol. Control*, vol. 49, no. 4, pp. 583–607, 2020.

[11] A. Singh *et al.*, "Joint encryption and compression-based watermarking technique for security of digital documents," *ACM Trans. Internet Technol. (TOIT)*, vol. 21, no. 1, pp. 1–20, 2021.

[12] J. Latham *et al.*, "Effect of scan pattern on complete-arch scans with 4 digital scanners," *J. Prosthet. Dent.*, vol. 123, no. 1, pp. 85–95, 2020.

[13] A. Nurunnabi, F. Teferle, J. Li, R. Lindenbergh, and A. Hunegnaw, "An efficient deep learning approach for ground point filtering in aerial laser scanning point clouds, *The Inter. Archives Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 31–38, 2021.

[14] M. Doneus, G. Mandlburger, and N. Doneus, "Archaeological ground point filtering of airborne laser scan derived point-clouds in a difficult mediterranean environment," *J. Comput. Appl. Archaeol.*, vol. 3, no. 1, pp. 92–108, 2020.

[15] G. Malaperdas, "Digitization in archival material conservation processes," *Eur. J. Eng. Technol. Res.*, vol. 6, no. 4, pp. 30–32, 2021.
[16] D. Zhang *et al.*, "A multi-label classification method using a hierarchical and transparent representation for paper-reviewer recommendation," *ACM Trans. Inf. Syst.*, vol. 38, no. 1, pp. 1–20, 2020.
[17] T. Hegghammer, "OCR with tesseract, amazon textract, and google document AI: A benchmarking experiment," *J. Comput. Soc. Sci.*, vol. 5, no. 1, pp. 861–882, 2022.
[18] M. Revilla-León, M. Sadeghpour, and M. Özcan, "An update on applications of 3D printing technologies used for processing polymers used in implant dentistry," *Odontology*, vol. 108, no. 3, pp. 331–338, 2020.
[19] C. Mangano, F. Luongo, M. Migliario, C. Mortellaro, and F. Mangano, "Combining intraoral scans, cone beam computed tomography and face scans: The virtual patient," *J. Craniofac. Surg.*, vol. 29, no. 8, pp. 2241–2246, 2018.
[20] A. Haleem and M. Javaid, "3D scanning applications in medical field: A literature-based review," *Clin. Epidemiol. Glob. Health*, vol. 7, no. 2, pp. 199–210, 2019.
[21] W. Hassan, Y. Yusoff, and N. Mardi, "Comparison of reconstructed rapid prototyping models produced by 3-dimensional printing and conventional stone models with different degrees of crowding," *Am J. Orthod. Dentofac.*, vol. 151, no. 1, pp. 209–218, 2017.
[22] K. Adam, A. Baig, S. Al-Maadeed, A. Bouridane, and S. El-Menshawy, "KERTAS: Dataset for automatic dating of ancient Arabic manuscripts," *Int. J. Doc. Anal. Recognit*, vol. 21, no. 4, pp. 283–290, 2018.
[23] X. Liu *et al.*, "Automatic organ segmentation for CT scans based on super-pixel and convolutional neural networks," *J. Digit. Imaging*, vol. 31, no. 5, pp. 748–760, 2018.
[24] C. Ben Rabah, G. Coatrieux, and R. Abdelfattah, "Automatic source scanner identification using 1D convolutional neural network," *Multimed. Tools Appl.*, vol. 81, no. 16, pp. 22789–22806, 2022.
[25] W. Qin, R. Elanwar, and M. Betke, "Text and metadata extraction from scanned arabic documents using support vector machines," *J. Inf. Sci.*, vol. 48, no. 2, pp. 268–279, 2022.
[26] A. Naïve and J. Barbosa, "Efficient accreditation document classification using naïve bayes classifier," *Indian J. Sci. Technol.*, vol. 15, no. 1, pp. 9–18, 2022.
[27] A. Kumar *et al.*, "Closing the loop: Automatically identifying abnormal imaging results in scanned documents," *J. Am. Med. Inf. Assoc.*, vol. 29, no. 5, pp. 831–840, 2022.
[28] F. Rashid, S. Gargaare, A. Aden, and A. Abdi, "Machine learning algorithms for document classification: Comparative analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 4, pp. 260–265, 2022.
[29] A. Guha, A. Alahmadi, D. Samanta, M. Khan, and A. Alahmadi, "A multi-modal approach to digital document stream segmentation for title insurance domain," *IEEE Access*, vol. 10, no. 1, pp. 11341–11353, 2022.
[30] S. Shreesha and H. Anita, "Classification of north and south Indian handwritten scripts using gabor wavelet features," *Indian J. Sci. Technol.*, vol. 15, no. 16, pp. 712–717, 2022.
[31] B. Omarov *et al.*, "Artificial intelligence in medicine: Real time electronic stethoscope for heart diseases detection," *Comput. Mater. Contin.*, vol. 70, no. 2, pp. 2815–2833, 2022.
[32] E. Hsu, I. Malagaris, Y. Kuo, R. Sultana, and K. Roberts, "Deep learning-based NLP data pipeline for EHR-scanned document information extraction," *JAMIA Open*, vol. 5, no. 2, pp. pp 1–12, 2022.
[33] G. BinMakhashen and S. Mahmoud, "Historical document layout analysis using anisotropic diffusion and geometric features," *Int. J. Digit. Libr.*, vol. 21, no. 3, pp. 329–342, 2020.
[34] S. Darwish and H. ELgohary, "Building an expert system for printer forensics: A new printer identification model based on niching genetic algorithm," *Expert Syst.*, vol. 38, no. 2, pp. 1–14, 2021.
[35] K. Aderghal, K. Afdel, J. Benois-Pineau, and G. Catheline, "Improving Alzheimer's stage categorization with convolutional neural network using transfer learning and different magnetic resonance imaging modalities," *Heliyon*, vol. 6, no. 12, pp. 1–13, 2020.
[36] K. Han *et al.*, "An expert-in-the-loop method for domain-specific document categorization based on small training data," *J. Assoc Inf. Sci. Technol.*, vol. 74, no. 6, pp. 669–684, 2023.

[37] S. Khaleefah, S. Mostafa, A. Mustapha, and M. Nasrudin, "The ideal effect of gabor filters and uniform local binary pattern combinations on deformed scanned paper images," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 33, no. 10, pp. 1219–1230, 2021.

[38] A. Jadli, M. Hain, and A. Hasbaoui, "An improved document image classification using deep transfer learning and feature reduction," *Int. J.*, vol. 10, no. 2, pp. 549–557, 2021.

[39] G. Nagy, "Twenty years of document image analysis in PAMI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 38–62, 2000.

[40] R. Gomez *et al.*, "ICDAR2017 robust reading challenge on COCO-Text," in *Proc. 2017 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Kyoto, Japan, 2018, pp. 1435–1443.

[41] K. Hashmi, A. Pagani, M. Liwicki, D. Stricker, and M. Afzal, "Cascade network with deformable composite backbone for formula detection in scanned document images," *Appl. Sci.*, vol. 11, no. 16, pp. 7610, 2021.

[42] J. Younas *et al.*, "Fi-Fo detector: Figure and formula detection using deformable networks," *Appl. Sci.*, vol. 10, no. 18, pp. 6460, 2020.

[43] L. Gao, X. Yi, Z. Jiang, L. Hao, and Z. Tang, "ICDAR2017 competition on page object detection," in *Proc. 2017 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Kyoto, Japan, 2017, pp. 1417–1422.

[44] X. Li, F. Yin, and C. Liu, "Page object detection from pdf document images by deep structured prediction and supervised clustering," in *Proc. 2018 24th Int. Conf. Pattern Recognit. (ICPR)*, Beijing, China, Aug. 20–24, 2018, pp. 3627–3632.

[45] N. Ranjan, "Document classification using LSTM neural network," *J. Data Min. Manag.*, vol. 2, no. 2, pp. 1–9, 2017.