information

MDPI

Article

Classification of English Words into Grammatical Notations Using Deep Learning Technique

Muhammad Imran [1,2,*], Sajjad Hussain Qureshi [3,*], Abrar Hussain Qureshi [4] and Norah Almusharraf [1]

1 Education Research Lab, Prince Sultan University, Riyadh 11586, Saudi Arabia; nmusharraf@psu.edu.sa
2 Department of English Language and Literature, Khazar University, Baku AZ1096, Azerbaijan
3 Department of Information Technology, Islamia University, Bahawalpur 06314, Pakistan
4 Department of English, University of Sahiwal, Sahiwal 57040, Pakistan; abrarqureshi@uosahiwal.edu.pk
* Correspondence: mimran@psu.edu.sa (M.I.); sajjads2002@yahoo.com (S.H.Q.)

Abstract: The impact of artificial intelligence (AI) on English language learning has become the center of attention in the past few decades. This study, with its potential to transform English language instruction and offer various instructional approaches, provides valuable insights and knowledge. To fully grasp the potential advantages of AI, more research is needed to improve, validate, and test AI algorithms and architectures. Grammatical notations provide a word's information to the readers. If a word's images are properly extracted and categorized using a CNN, it can help non-native English speakers improve their learning habits. The classification of parts of speech into different grammatical notations is the major problem that non-native English learners face. This situation stresses the need to develop a computer-based system using a machine learning algorithm to classify words into proper grammatical notations. A convolutional neural network (CNN) was applied to classify English words into nine classes: noun, pronoun, adjective, determiner, verb, adverb, preposition, conjunction, and interjection. A simulation of the selected model was performed in MATLAB. The model achieved an overall accuracy of 97.22%. The CNN showed 100% accuracy for pronouns, determiners, verbs, adverbs, and prepositions; 95% for nouns, adjectives, and conjunctions; and 90% for interjections. The significant results ($p < 0.0001$) of the chi-square test supported the use of the CNN by non-native English learners. The proposed approach is an important source of word classification for non-native English learners by putting the word image into the model. This not only helps beginners in English learning but also helps in setting standards for evaluating documents.

Keywords: artificial intelligence; grammatical notations; CNN; deep learning

Citation: Imran, M.; Qureshi, S.H.; Qureshi, A.H.; Almusharraf, N. Classification of English Words into Grammatical Notations Using Deep Learning Technique. *Information* **2024**, *15*, 801. https://doi.org/10.3390/info15120801

Academic Editor: Ismailcem Budak Arpinar

Received: 6 October 2024
Revised: 21 November 2024
Accepted: 4 December 2024
Published: 11 December 2024

1. Introduction

Computer software provides learners with access to pertinent information quickly. Learners are motivated to learn more by using machine-based resources that offer diverse educational materials in a short time frame. Software alters language learning schema by allowing students to tailor their education to their interests. Additionally, machine-based models stimulate the learners' auditory and visual senses. The literature has also described the role of WhatsApp applications in English conversation magazines to enhance learners' vocabulary [1]. Based on these studies' findings, this application improved learners' vocabulary and word choice. The use of wiki technology plays a vital role in enhancing students' writing abilities [2]. The deployment of technology in the classroom as a teaching tool enhances the participation of students in class [3]. Technology-based tools enhance motivation and accessibility. Machine learning algorithms allow students to change their learning process and obtain a wealth of knowledge that is not available from their lecturers [4]. Regardless of the learner's background, software is also a source to increase awareness about computer technology, offering equal knowledge and skills. They make learners more confident, self-reliant, and goal-oriented. Most of this information is communicated in English [5]. The use of this information comes under the domain of Natural

Language Processing (NLP). NLP has profited enormously from the evolution of deep neural networks (DNNs) because they require less manual feature engineering. Document classification, text recognition, grammar checking, and parts of speech identification and classification come under the umbrella of NLP.

In the field of computer vision, digital document analysis is a longstanding activity that has gained momentum due to the use of digital cameras and cell phones in everyday life. Text content extraction from scanned or photographed documents is an essential step in digitization to allow for simple text searches and indexing [6]. Support vector machine (SVM) classifiers are thought to be among the best methods for classification problems, but they do not support large datasets [7]. Various deep learning approaches have been proposed to cover more difficult problems and yield high-quality results on varied tasks [8]. The primary purpose of the first convolutional neural network (CNN) was to categorize letters and digits [9]. Ref. [10] focused on document image classification without OCR results. Convolutions in CNNs are thought of as feature extractors. They can be set up to match particular features, like edges with shapes, or to reassess their internal weights in successive iterations in light of a particular context.

Neural network models have immense importance in the field of speech identification [11]. Neural network models use word vector representation and utilize this technique for word classification [12]. Efficient and direct models like linear classifiers are regularly utilized as strong baselines for sentence classification issues [13]. Regardless of their simplicity, they regularly obtain cutting-edge results if proper features are selected. However, convolutional or recurrent neural networks have achieved immense importance in NLP [14–16]. Each model has indicated incredible performance, frequently surpassing the forecast time and slightly enhancing the accuracy. Deep neural networks have two main types: convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [17]. The current NLP writing performance does not ensure that the intended goals are met. For instance, RNNs perform well on documented sentiment characterization [18,19], and it was demonstrated that gated CNNs perform well using long short-term memory (LSTM) on language processing assignments, even though LSTM has been viewed as more qualified for quite some time. CNNs (convolutional neural networks) have probably been the most persuasive advancements in PC vision. This system became the priority when reference [20] won the ImageNet competition by decreasing the number of classification mistakes from 26 to 15. CNNs take inspiration from the visual cortex. The visual cortex has little districts of cells that are definite areas of the visual field. This thought was developed by an interesting trial by Hubel and Wiesel [21]. They indicated that some individual neuron cells in the cerebrum reacted uniquely when the edges of a specific direction were seen. For instance, a few neurons terminated when presented with vertical edges, and some terminated when presented with flat or corner-to-corner edges. They discovered that these neurons were arranged in columnar structures and could enable visual discernment. This proposal of particular segments within a framework having explicit functions is one that machines also use and is the premise behind CNNs [9,22]. Currently, deep learning developments in text mining usually concentrate on developing increasingly intricate and sophisticated neural network-based models to manage large text or text image documents. The role of CNN-based models for sentence categorization has already been reported in the literature [23]. The researchers used character-level identification to make their model more efficient [24–26]. However, it requires more layers and the parameters require more training, resulting in huge computational costs. Considering the potential of CNNs and the requirement of the project, this study is performed to develop an architecture to classify English words and images into different parts of speech using a CNN.

2. Materials and Methods

Newspaper texts have extensively been studied in the past research [27–30]. However, this present research was conducted to develop the architecture of a neural network model that took images of English words through scanning and would help classify them into different parts of speech (Figure 1). Mostly, word-to-vector techniques are utilized to classify the words. However, we used image-based techniques for the development of the architecture. The details of the proposed model are as follows.

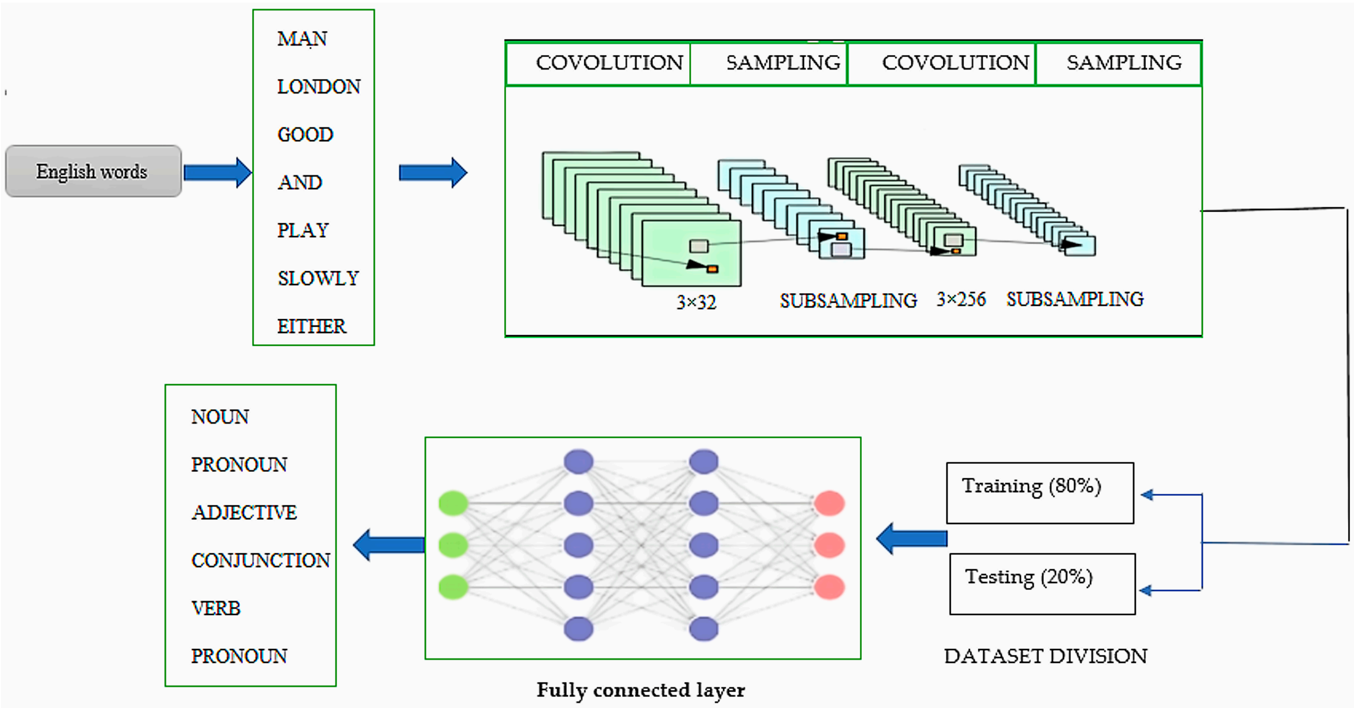


Figure 1. Conceptual design of the proposed model.

2.1. Dataset Preparation

In text image-based recognition research, appropriate datasets are required from training to testing stages. Therefore, various techniques can be employed to acquire the images of the text. The current study collected word images by scanning data relating to different parts of speech from different sources and labeled them with proper grammatical notation. Images had different resolutions and sizes. Their dimensions were properly adjusted by cropping the images and resizing them. Several sized photos in the dataset resulted from the fusion of the assembled dataset from various pre-existing photos. All images are reduced to a dimensional size of 224×224, as CNNs normally require input photos of the same size for accurate processing. The normalization process was applied to all scanned images. It guarantees that the features have a standard size and are centered. The training and testing datasets are maintained using the same mean and standard deviation values. Data augmentation was executed to produce more training data from different angles from the already existing datasets. Every input image can be moved randomly, rotated, zoomed in or out, flipped vertically or horizontally, and so on.

The CNN will gain knowledge from various training instances in this way. It will also help to overcome the overfitting problem of the model. Moreover, adding a dropout layer will play a vital role in regularization, which in turn is a solution to the model overfitting problem. A non-participatory approach was explored during data collection. A native English learner categorized the words into different grammatical notations to label the dataset. More preprocessing was required to ensure that the deep neural network produced consistent results. Images in the region of interest with the highest resolution are deemed suitable for inclusion in the dataset. Moreover, image parameters were adjusted to minimize the model training period.

2.2. Sample Division

Nine hundred word images (900) were collected. Later, the datasets were divided into training, validation, and test datasets. Training datasets were used to train the model, while validation datasets were used to tune hyperparameters. The model's actual results were observed through test datasets.

2.3. Convolutional Neural Network

The human brain served as the model for the convolutional neural network (CNN). Many machine learning features, such as feature maps, subsampling, and shared weights, are included. These machines have been frequently used in the past for the successful identification of handwritten materials. However, the introduction of neural networks, especially CNNs, has also made it possible to identify digits and text within the image.

Typically, a CNN model consists of an input, hidden, and output layer. The proposed model receives the input image of size 224 × 224 × 3 in the input layer. Hidden layers perform convolutional operations to minimize the image dimensions and retrieve the most relevant pattern. Fully connected layers use these patterns to predict the complex classes of word notations (Figure 2). The input layer acts as a door for the test data to be classified. The main functions of classification are performed in hidden layers. The model has a convolutional layer where feature extraction and detection take place. Several feature maps are produced by convolving the inputs of each convolutional layer with a unique set of kernels.

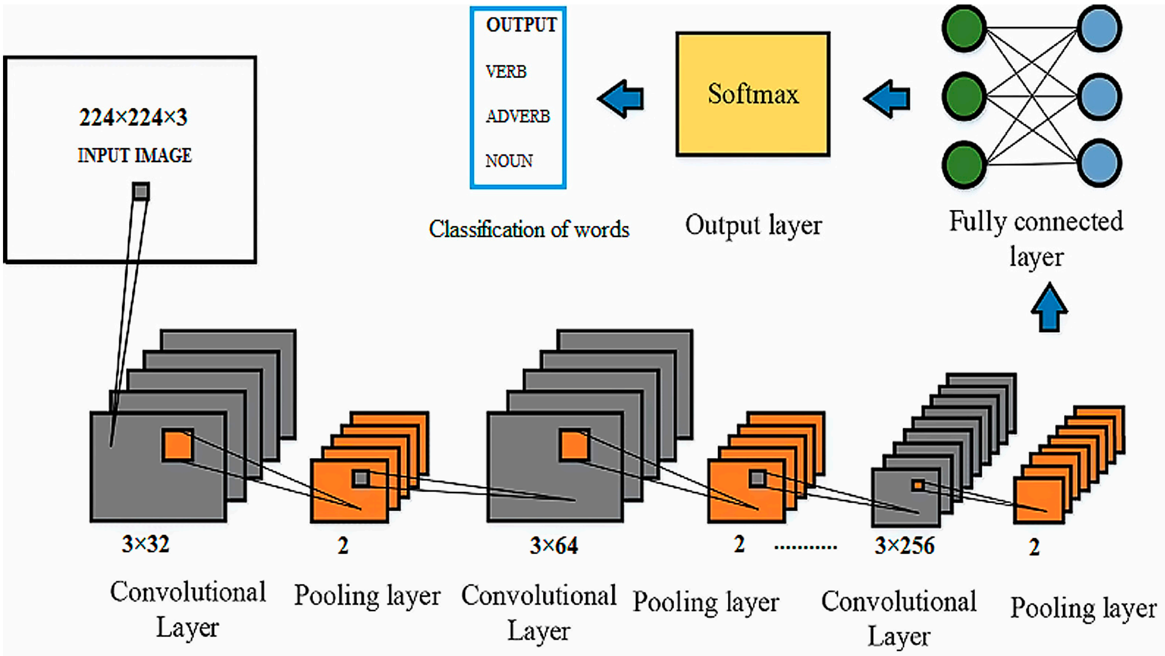


Figure 2. The proposed architecture of the lightweight model.

Activation functions introduce nonlinearity in deep learning models. An activation function is implemented following each convolutional layer in CNN models. A convolution layer often follows a pooling layer. This layer is important to identify the area of interest in the image. It is impossible to correctly identify where each feature is precisely located. Moreover, it helps to clean irrelevant data from the feature maps. This layer performs pooling using the mean or max pooling function. Max pooling has emerged as the most often used pooling function, which includes the maximum value selected from each region of the featured map (Figure 3). The current model also used the max pooling layer after each convolutional layer. In CNNs, the last layer comprises fully connected layers. It is also considered an output layer. Softmax classifier was used to classify the English words into different grammatical notations.

Figure 3. Feature mapping and max pooling technique of CNN model.

Each unit of this layer is completely connected to the previous layer, like other layers of CNNs.

The CNN architecture was created using MATLAB to divide the word into nine grammatical notations (Table 1). The suggested CNN model has five convolutional and four max pooling layers. The last layer is a fully connected layer after the last convolutional layer.

Table 1. Summary of the CNN layers.

Layer	Size	Other Parameters
Input	224 × 224 × 3	
Convolution 1	3 × 32	padding = same
Max Pooling	2	stride = 2
Convolution 2	3 × 64	padding = same
Max Pooling	2	stride = 2
Convolution 3	3 × 128	padding = same
Max Pooling	2	stride = 2
Convolution 4	3 × 256	padding = same
Max Pooling	2	stride = 2
Convolution 5	3 × 256	padding = same
Fully Connected + Softmax	9	

The input layer received the image of 224 × 224 × 3 size. The ReLU layer was used as a batch normalization layer between convolutional and nonlinear layers to expedite neural network training and lessen susceptibility to initialization. The first convolutional layer contained a kernel size of 3 × 3 with 32 filters. It was followed by a max pooling layer with a size of 2 × 2. The second convolutional layer contained a kernel size of 3 × 3 with 64 filters. It was followed by a max pooling layer with a size of 2 × 2. The third convolutional layer contains a kernel size of 3 × 3 with 128 filters. A max pooling layer with a stride of 2 follows it. The fourth and fifth convolutional layers contained a kernel size of 3 × 3 with 256 filters. It is followed by a max pooling layer with a size of 2 × 2. After the last convolutional layer, the OutputSize parameter was set equal to the whole number of classes in the target data. The output size in this case was 09, which was equivalent to nine classes. The output of the fully connected layer was normalized by the Softmax activation function. A stochastic gradient with a learning rate of 0.00001 was used to train the model. Twelve epochs were selected to adjust the network weight and bias, which helped to reduce the error between actual and predicted classes.

The function of the proposed classifier is as follows.

1. Input: The input consists of N-word images with K classification labels, which we refer to as training sets.

2. Learning: Each precise class membership is described using the training set in this stage. In essence, this stage is known as learning or training the classifier.
3. Assessment: In this stage, the classifier quality to guess the classification labels of previously unseen word images. We create a connection between the classifier's estimated labels and the actual labels on the word images.

2.4. Training of Model

The model was trained using word images. The model neurons acting on the word images produced the intended result for the training session. During the training session, iterations were carried out to minimize the error rate and modify the input weights.

2.5. Validation and Testing of the Model

In order to confirm that the model's actual outputs match the intended outputs, model training was continued. A validation process was performed to fine-tune the neuron weights. The model's testing was performed to compare model results with actual results.

2.6. Research Quality

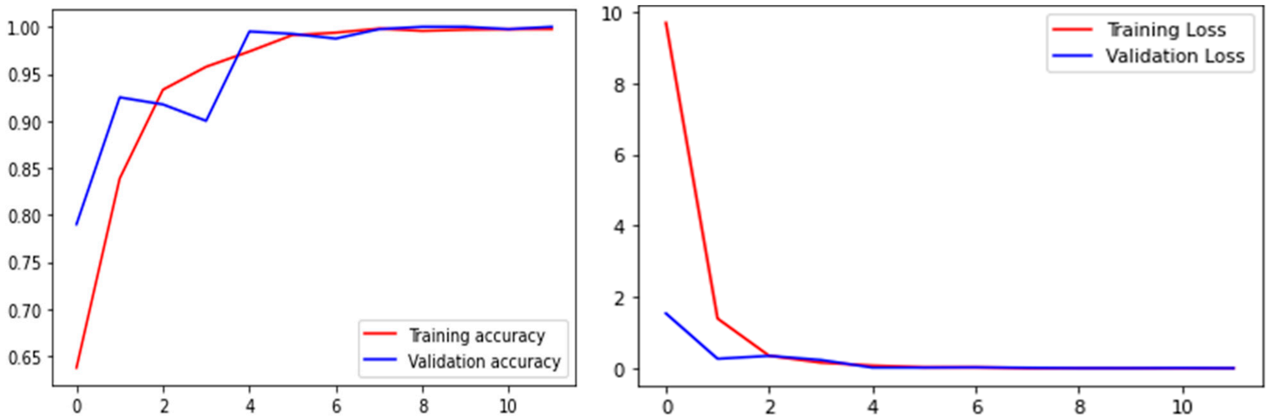
In [31], the authors developed statistical measures to judge the accuracy of the model. These measures were used to assess the prediction of the proposed model. The accuracy of the model was measured using the following formula:

$$\text{Accuracy(A)} = (\text{Tp} + \text{Tn})/(\text{Tp} + \text{Tn} + \text{Fp} + \text{Fn})$$

3. Results

The conventional method of learning grammatical notation of English for non-native English learners is the use of manual dictionaries. This study's objective was to facilitate learners by using the image of the English words as a source for the model to describe it in its proper notation. A total of hundred (100) images of each grammatical notation were taken. A total of 80% of the images were used to train the model, while 20% were used to test the model's accuracy. Figure 4a,b provide a graphical depiction of the relationship between training and validation parameters on word images. Figure 4a,b display the accuracy and loss measurements during the training and validation of the suggested model, respectively. The twelve (12) epochs were set to check the training and validation accuracy of the model. At the start, it was observed that the model training accuracy increased gradually with an increase in its validation. The training accuracy was 93%, while its validation accuracy was 90% at epoch 2. The training accuracy and validation accuracy was 95% at epoch 4. The model continuously increased its learning. Its training accuracy was 98% at epoch 6, with a validation accuracy of 96%. At epochs 8 to 12, the training and validation accuracy remained the same (98%) (Figure 4a).

The CNN architecture elaborated in this paper provides more reliable feedback to the learners in a short time frame. The model's authenticity was checked through its accuracy. It was observed that the model showed 95% accuracy for nouns, 100% for pronouns, 95% for adjectives, 100% for determiners, 100% for verbs, 100% for adverbs, 100% for prepositions, 95% for conjunctions, and 90% for interjection (Table 2). The overall accuracy of the model was 97.22%. It was further observed that 5% of the nouns were incorrectly judged as verbs, 5% of adjectives were incorrectly judged as verbs, 5% of conjunctions were incorrectly judged as interjections, and 10% of interjections were incorrectly judged as conjunctions (Figure 5). Later on, the model was checked using test data. It was observed that the model correctly identified adjectives, adverbs, nouns, and verbs (Figure 6).



(**a**) (**b**)

Figure 4. (**a**) Training and validation accuracy of the proposed model. (**b**) Training and validation loss of the proposed model.

Table 2. Accuracy of parts of speech into nine classes.

Parts of Speech	Total	Classification		Accuracy (%)
		Correctly	Incorrectly	
Noun	20	19	1	95
Pronoun	20	20	0	100
Adjective	20	19	1	95
Determiner	20	20	0	100
Verb	20	20	0	100
Adverb	20	20	0	100
Preposition	20	20	0	100
Conjunction	20	19	1	95
Interjection	20	18	2	90
Average				97.22

A comparison of the identification of grammatical notations between native English learners (standard), non-native English learners, and the CNN was performed (Figure 7). Twenty words of each notation were fed to each category. Native English learners' answers are used as a check for all classes of grammatical notations. In the case of nouns, non-native English learners correctly recognized fifteen (15) words, while the CNN identified nineteen (19) words. In the case of pronouns, non-native English learners correctly recognized twelve (12) words, while the CNN identified twenty (20) words. In the case of adjectives, non-native English learners correctly recognized ten (10) words, while the CNN identified nineteen (19) words. In the case of determiner, non-native English learners correctly recognized eight (08) words, while the CNN identified twenty (20) words. Regarding verbs, non-native English learners correctly recognized sixteen (16) words, while the CNN identified twenty (20). In the case of adverbs, non-native English learners correctly recognized twelve (12) words, while the CNN identified twenty (20) words. In the case of prepositions, non-native English learners correctly recognized eighteen (18) words, while the CNN identified twenty (20) words. In the case of conjunction, non-native English learners correctly recognized twelve (12) words, while the CNN identified nineteen (19) words. Similarly, non-native English learners correctly recognized eleven (11) words, while the CNN correctly identified eighteen (18) words.

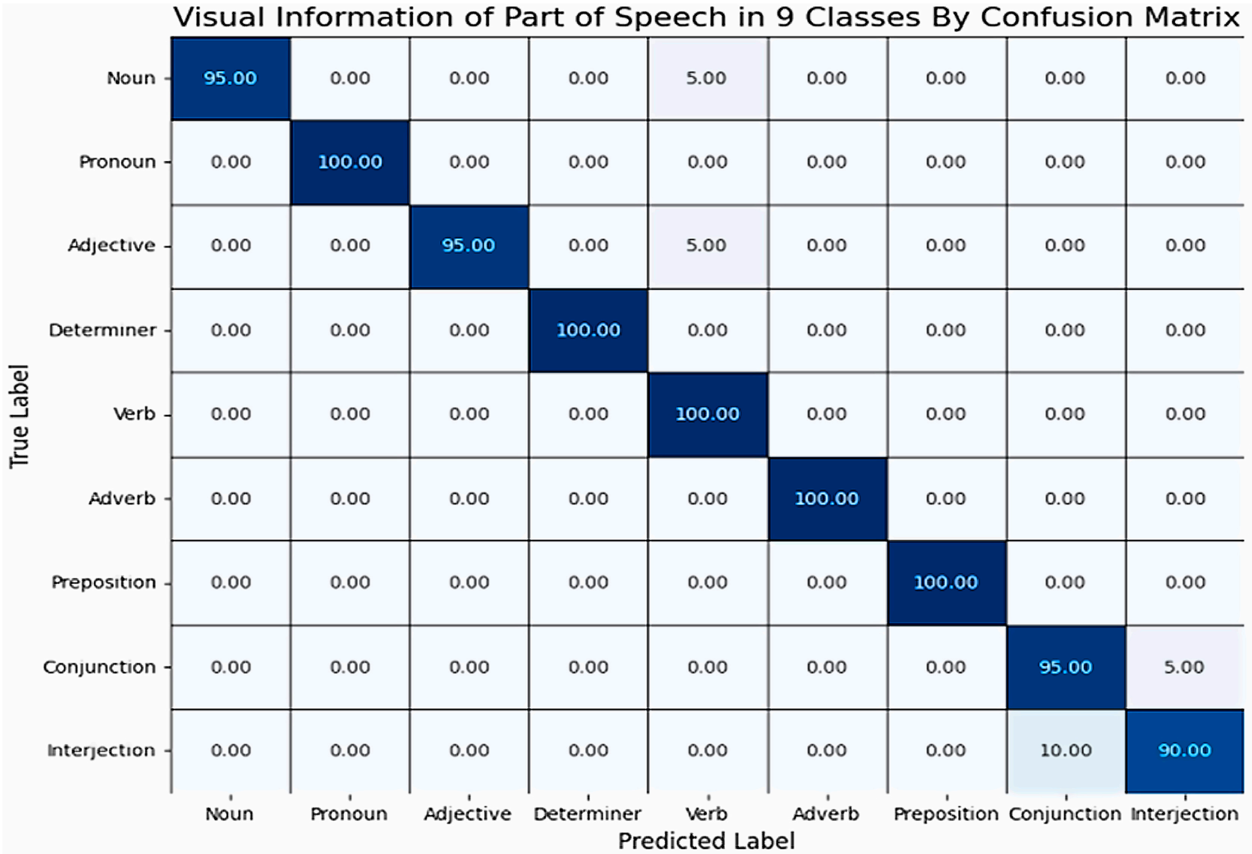


Figure 5. Confusion Matrix of parts of speech into nine classes.

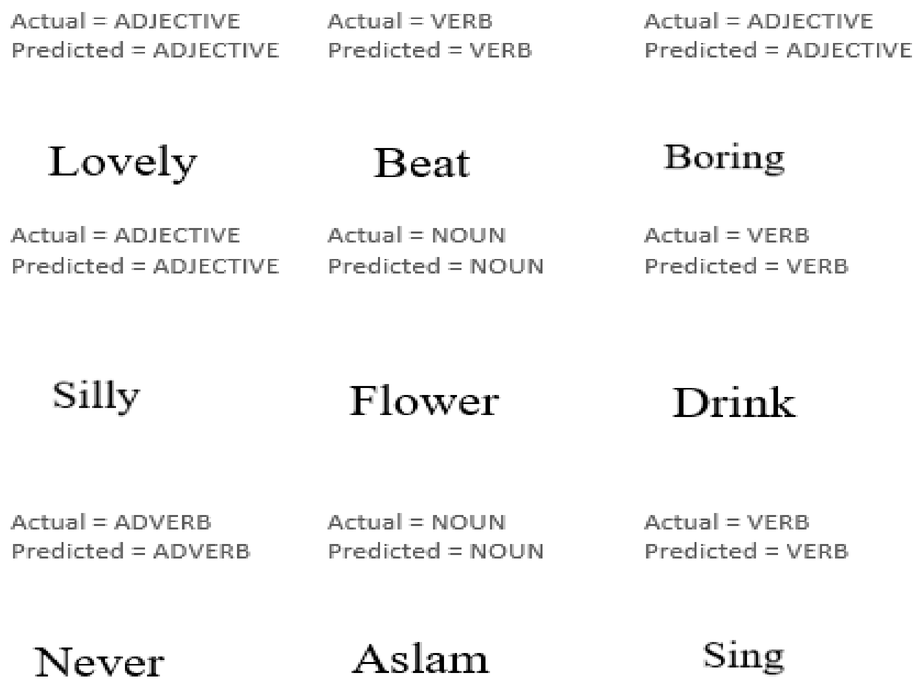


Figure 6. Results of model predictions.

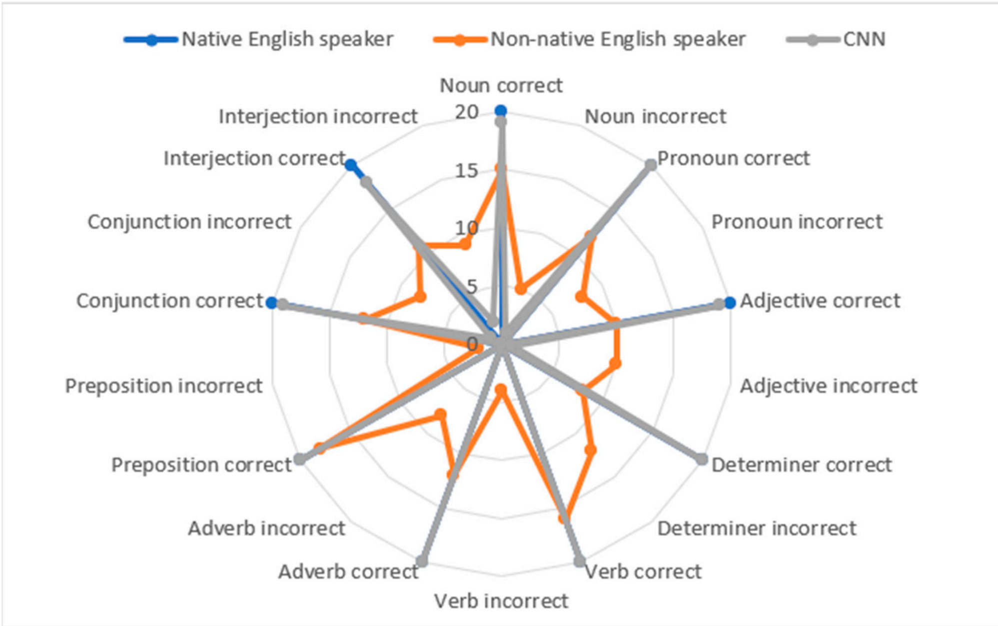


Figure 7. Comparison of grammatical notation recognition.

A chi-square test was also performed to check the significance of differences among the studied classes (Table 3). A significant difference was exhibited among the classes at $p < 0.05$. The test showed 127.59 and 0.0001 for the chi-square test and the p value, respectively.

Table 3. Comparison of native and non-native English learners with CNN.

Class	Words Classification		Chi-Square Value	p Value
	Correct	Incorrect		
Native English Learner (check)	180	0	127.59	0.0001 *
Non-Native English Learner	114	66		
CNN	175	05		

CNN = convolutional neural network, * significant at $p < 0.05$.

4. Discussions

Non-English natives use different AI technologies, like ChatGPT, Perplexity, and OpenAI, to learn English vocabulary. ChatGPT is a recent innovation that is described as a massive auto-completion engine. ChatGPT has spurred fresh research on the use of technology in learning and writing about different aspects of human needs. AI is able to respond to each learner's learning style, learning limitations, and strengths through sophisticated data filtering [32]. Using this information, artificial intelligence can modify the pace, challenge, and learning materials to meet the needs of each learner better and increase learning effectiveness. Through this, it will be possible to lessen the possible detrimental effects of AI on motivation while boosting its beneficial effects. AI-based tools for language learning offer English language learners individualized feedback, prompt responses, and detailed interaction [33]. AI is incorporated more deeply into learning platforms, and there are concerns over how it may affect conventional teaching strategies, student autonomy, and the validity of learning. The more robust, accurate, and validated architecture can answer it best. Furthermore, based on the procedure and methodology, the learners for this investigation were classified as beginners based on their self-reported lack of prior exposure to formal English education and the absence of standardized proficiency assessments. While this classification provides a general context, the lack of standardized measures is acknowledged as a limitation in this comparison.

The CNN worked well with large datasets, extracted desirable features automatically, and marked quick predictions. It excelled at identifying basic textual patterns. The minimum numbers of layers were utilized to minimize the computational cost. In the case of the recurrent neural network (RNN), the process of training a model is challenging and time-consuming. It frequently takes a long time to find nonlinearities in the data and is vulnerable to the vanishing gradient issue [34]. Similarly, the long short-term memory (LSTM) model increases the problem's dimensionality and complicates the search for the best solution by employing backpropagation, which adds cost and complexity [35–37].

This study obtained ideas from the works concerning the use of CNNs for document classification [38,39]. The authors in [40] primarily used CNNs for text classification. They elaborated that CNN layers have the capability to retain the hierarchical record to obtain information about undocumented text. This methodology figured out the similarities among the documents. The authors in [41] described the importance of linear models in word representation learning. Refs. [42,43] elaborated that these models could be trained on large datasets in a short time. Their performance is the same as the conventional method, but the training time is much lower as compared to the state-of-the-art models [38]. The presented RVL-CDIP dataset gives an enormous dataset to record classification and permits the training of DNN from scratch, providing considerable results over hand-created options. This study further elaborated that the availability of a sufficient amount of data and region-specific learning is not necessary. CNN models have consequently been demonstrated to be successful in natural language processing (NLP) and have accomplished amazing outcomes in semantic parsing [44] and sentence displaying [45]. The current study achieved 97.22% accuracy compared to 91.13% achieved by [39]. Our work is insightful, like [46], who demonstrated that pre-trained neural networks performed more accurately in image classification than the other models. Picture descriptor algorithms demonstrate more accurate results, showing more credible outcomes. The difficulties are faced when there are low inter-class and high intra-class variations [47,48]. Ref. [39] demonstrated an extraordinary improvement in precision by applying deep models and transferring learning from the domain of real-world pictures to the domain of documented pictures. This makes it possible to use DNN architecture with limited numbers of training data. In short, the CNN produces profoundly effective compact descriptions in the classification of documents to a great extent, outperforming prior SIFT-based encoding plans [49]. The significant results from the chi-square test favor using CNNs to improve the knowledge of non-native learners through machine learning algorithms.

MobileNet engineering [50] is an efficient model improved for versatile and implanted vision applications. MobileNets are a reaction to the pattern that, as of late, made neural networks consistently more profound and more muddled to accomplish higher precision. Ref. [51] explored CNNs and fundamental RNNs to describe their classification pattern. They reported that CNNs provided more reliable information than RNNs. The RNN processed a weighted blend of all words in the sentence while the CNN separated the most enlightening ngrams for the relation and only considered their subsequent activation [44]. Moreover, Ref. [52] described the importance of CNNs over GRU/LSTM for the classification of long sentences. Moreover, Ref. [53] accomplished a better execution of the CNN than LSTM for answer determination. The authors in [19] further contended that CNNs could show long-term reliance and achieve new standards for language modeling.

5. Conclusions

The findings of this study help non-native English learners identify the issues they were facing in classifying English words into grammatical notations. They further elaborated on the difficulty of separating words into their grammatical notations. These problems increased their learning time due to the need to manually search for the words in a dictionary. This study used the power of a computer to execute a machine-learning algorithm. It is a tool that learners can utilize because it assists them in resolving their educational issues and figuring out how to apply what they have learned in practical and significant ways. This research also revealed that the use of technology contributes significantly to language learning at the student's own pace, aids in self-understanding, does not need any coaching, and greatly increases motivation for language learning. Future research will involve more rigorous linguistic proficiency assessments to provide a detailed basis for comparison and further validate the model's application across different learner profiles. Additionally, this study recommends that learners use the power of deep learning to improve their language proficiency since it fosters creativity and offers engaging, enjoyable, and innovative language learning opportunities. This study not only provides the solution to the problems of non-native English learners but also provides a direction for deploying the CNN architecture on mobile phones to achieve more fruitful results.

6. Limitations of This Study

This study has certain limitations that warrant consideration. First, the non-native English learners involved were classified as beginners based on general observations of their educational background and rural upbringing without formal proficiency assessments. This lack of standardized linguistic proficiency metrics limits the precision of the comparison between human learners and the CNN model. Additionally, this study focuses on a specific context with limited generalizability to more diverse learner profiles or advanced proficiency levels. Future research should incorporate standardized tests to assess linguistic proficiency and explore the model's applicability across varied educational and cultural settings.

Author Contributions: Conceptualization, A.H.Q.; methodology, M.I.; validation, S.H.Q.; formal analysis, S.H.Q.; writing and original draft preparation, A.H.Q. and M.I.; writing, review and editing, N.A.; visualization, S.H.Q.; supervision, A.H.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This study is funded by Prince Sultan University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data relevant to the research are available in the paper.

Acknowledgments: The authors would like to thank Prince Sultan University for its financial (APC) and technical support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Afsyah, S. WhatsApp Application in English Language Teaching (ELT) Context: Media to Describe People. *Utamax J. Ultim. Res. Trends Educ.* **2019**, *1*, 23–28. [CrossRef]
2. Rahimi, M.; Fathi, J. Exploring the Impact of Wiki-Mediated Collaborative Writing on EFL Students' Writing Performance, Writing Self-Regulation, and Writing Self-Efficacy: A Mixed Methods Study. *Comput. Assist. Lang. Learn* **2022**, *35*, 2627–2674. [CrossRef]
3. Hasas, A.; Sadaat, S.N.; Hakimi, M.; Quchi, M.M. Interactive Learning in Afghanistan: Feasibility of Implementing IoT Connected Devices in Classrooms. *Am. J. Smart Technol. Solut.* **2024**, *3*, 8–16. [CrossRef]
4. Luo, Y.; Han, X.; Zhang, C. Prediction of Learning Outcomes with a Machine Learning Algorithm Based on Online Learning Behavior Data in Blended Courses. *Asia Pac. Educ. Rev.* **2024**, *25*, 267–285. [CrossRef]
5. Reynolds, B.L.; Yu, M.H.; Ha, X. Van Administrative Staff Learning and Using English to Communicate with International Students. *Innov. Lang. Learn. Teach.* **2023**, *17*, 424–439. [CrossRef]
6. Zhou, Y.; Zhang, Z.; Wang, X.; Sheng, Q.; Zhao, R. Multimodal archive resources organization based on deep learning: A prospective framework. *Aslib J. Inf. Manag.* **2024**. [CrossRef]
7. Patil, S.S.; Mareeswari, V.; Chaitra, V.; Singh, P. Recognition of Handwritten Digits with the Help of Deep Learning. In *Lecture Notes in Networks and Systems*; Springer: Singapore, 2020; Volume 107. [CrossRef]
8. Cao, J.; Li, Y.; Sun, H.; Xie, J.; Huang, K.; Pang, Y. A survey on deep learning based visual object detection. *J. Image Graph.* **2022**, *27*, 1697–1722. [CrossRef]
9. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]

10. Chen, N.; Blostein, D. A survey of document image classification: Problem statement, classifier architecture and performance evaluation. *Int. J. Doc. Anal. Recognit.* **2007**, *10*, 1–16. [CrossRef]
11. Mahfood, B.; Elnagar, A.; Kamalov, F. Emotion Recognition from Speech Using Convolutional Neural Networks. In *Lecture Notes in Networks and Systems*; Springer: Singapore, 2023; Volume 572.
12. Mutinda, J.; Mwangi, W.; Okeyo, G. Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network. *Appl. Sci.* **2023**, *13*, 1445. [CrossRef]
13. Ning, B.; Zhao, D.; Zhang, X.; Wang, C.; Song, S. UMP-MG: A Uni-Directed Message-Passing Multi-Label Generation Model for Hierarchical Text Classification. *Data Sci. Eng.* **2023**, *8*, 112–123. [CrossRef]
14. Magalhães, D.; Lima, R.H.R.; Pozo, A. Creating Deep Neural Networks for Text Classification Tasks Using Grammar Genetic Programming[Formula Presented]. *Appl. Soft Comput.* **2023**, *135*, 110009. [CrossRef]
15. Tan, C.; Ren, Y.; Wang, C. An Adaptive Convolution with Label Embedding for Text Classification. *Appl. Intell.* **2023**, *53*, 804–812. [CrossRef]
16. Agbesi, V.K.; Chen, W.; Ameneshewa, A.S.; Odame, E.; Dumor, K.; Browne, J.A. Efficient Adaptive Convolutional Model Based on Label Embedding for Text Classification Using Low Resource Languages. In Proceedings of the 2023 7th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, Kuala Lumpur, Malaysia, 23–24 April 2023.
17. Hasbullah, S.; Mohd Zahid, M.S.; Mandala, S. Detection of Myocardial Infarction Using Hybrid Models of Convolutional Neural Network and Recurrent Neural Network. *BioMedInformatics* **2023**, *3*, 33. [CrossRef]
18. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for Target-Dependent Sentiment Classification. In Proceedings of the COLING 2016—26th International Conference on Computational Linguistics, Osaka, Japan, 11–16 December 2016.
19. Dauphin, Y.N.; Fan, A.; Auli, M.; Grangier, D. Language Modeling with Gated Convolutional Networks. In Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, Australia, 6–11 August 2017; Volume 2.
20. Krizhevsky, A.; Sutskever, I.H. Imagenet Classification with Deep Convolutional Neural Network. In *Advances in Neural Information Processing Systems*; Elsevier Ltd.: Amsterdam, The Netherlands, 2012; pp. 1097–1105.
21. Hubel, D.H.; Wiesel, T.N. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *J. Physiol.* **1962**, *160*, 106. [CrossRef]
22. Imran, M.; Almusharraf, N. Digital Learning Demand and Applicability of Quality 4.0 for Future Education: A Systematic Review. *Int. J. Eng. Pedagog.* **2024**, *14*, 38. [CrossRef]
23. Rezaeenour, J.; Ahmadi, M.; Jelodar, H.; Shahrooei, R. Systematic review of content analysis algorithms based on deep neural networks. *Multimed. Tools Appl.* **2023**, *82*, 17879–17903. [CrossRef]
24. Chauhan, V.K.; Singh, S.; Sharma, A. HCR-Net: A deep learning based script independent handwritten character recognition network. *Multimed. Tools Appl.* **2024**, *83*, 78433–78467. [CrossRef]
25. Hemanth, G.R.; Jayasree, M.; Venii, S.K.; Akshaya, P.; Saranya, R. CNN-RNN Based Handwritten Text Recognition. *Ictact J. Soft Comput.* **2021**, *12*, 2457–2463.
26. Prabakaran, N.; Kannadasan, R.; Krishnamoorthy, A.; Kakani, V. A Bidirectional LSTM approach for written script auto evaluation using keywords-based pattern matching. *Nat. Lang. Process. J.* **2023**, *5*, 100033. [CrossRef]
27. Ahmad, M.; Mahmood, M.A.; Mahmood, M.I.; Siddique, A.R. Use of modal verbs as stance markers in Pakistani English newspaper editorials. *Online J. Commun. Media Technol.* **2019**, *9*, e201903. [CrossRef]
28. Ahmad, M.; Mahmood, M.A.; Farukh, A. Use of modals as stance markers: A corpus-based study on Pakistani English newspaper editorials. *Asia Pac. Media Educ.* **2020**, *30*, 108–125. [CrossRef]
29. Buriro, W.M.; Zaki, S.; Ashraf, A. September. Grammatical features of Pakistani English: A Corpus Analysis of Pakistani English Newspapers and Magazines. *Linguist. Forum-A J. Linguist.* **2023**, *5*, 89–113.
30. Fatima, M.; Siddique, A.R.; Ahmad, M.; Mahmood, M.A. Exploring linguistic variation in Pakistani English newspaper editorials through multidimensional analysis. *Newsp. Res. J.* **2023**, *44*, 425–451. [CrossRef]
31. Muthukannan, K.; Latha, P. Fuzzy inference system based unhealthy region classification in plant leaf image. *Int. J. Comput. Inf. Eng.* **2015**, *8*, 2103–2107.
32. Maqbool, M.A.; Asif, M.; Imran, M.; Bibi, S.; Almusharraf, N. Emerging e-learning trends: A study of faculty perceptions and impact of collaborative techniques using fuzzy interface system. *Soc. Sci. Humanit. Open* **2024**, *10*, 101035. [CrossRef]
33. Imran, M.; Almusharraf, N.; Ahmed, S.; Mansoor, M.I. Personalization of E-Learning: Future Trends, Opportunities, and Challenges. *Int. J. Interact. Mob. Technol.* **2024**, *18*, 4. [CrossRef]
34. Malhotra, R.; Singh, P. Recent advances in deep learning models: A systematic literature review. *Multimed. Tools Appl.* **2023**, *82*, 44977–45060. [CrossRef]
35. Al-Selwi, S.M.; Hassan, M.F.; Abdulkadir, S.J.; Muneer, A. LSTM Inefficiency in Long-Term Dependencies Regression Problems. *J. Adv. Res. Appl. Sci. Eng. Technol.* **2023**, *30*, 16–31. [CrossRef]
36. El-Shafai, W.; Ali, A.M.; El-Nabi, S.A.; El-Rabaie ES, M.; Abd El-Samie, F.E. Single image super-resolution approaches in medical images based-deep learning: A survey. *Multimed. Tools Appl.* **2024**, *83*, 30467–30503. [CrossRef]
37. Al Saleh, R.; Driss, M.; Almomani, I. CBiLSTM: A hybrid deep learning model for efficient reputation assessment of cloud services. *IEEE Access* **2022**, *10*, 35321–35335. [CrossRef]
38. Harley, J.M.; Bouchet, F.; Hussain, M.S.; Azevedo, R.; Calvo, R. A Multi-Componential Analysis of Emotions during Complex Learning with an Intelligent Multi-Agent System. *Comput. Human Behav.* **2015**, *48*, 615–625. [CrossRef]

39. Afzal, M.Z.; Kolsch, A.; Ahmed, S.; Liwicki, M. Cutting the Error by Half: Investigation of Very Deep CNN and Advanced Training Strategies for Document Image Classification. In Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, Kyoto, Japan, 9–15 November 2017; Volume 1.
40. Kang, L.; Ye, P.; Li, Y.; Doermann, D. Convolutional neural networks for no-reference image quality assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1733–1740.
41. Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of Tricks for Efficient Text Classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017—Proceedings of Conference, Valencia, Spain, 3–7 April 2017; Volume 2.
42. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed Representations of Words and Phrases and Their Compositionality. In Proceedings of the Conference on Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013.
43. Levy, O.; Goldberg, Y.; Dagan, I. Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Trans. Assoc. Comput. Linguist.* **2015**, *3*, 211–225. [CrossRef]
44. Yih, W.T.; He, X.; Meek, C. Semantic Parsing for Single-Relation Question Answering. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014—Proceedings of the Conference, Baltimore, MD, USA, 22–27 June 2014; Volume 2.
45. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A Convolutional Neural Network for Modelling Sentences. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014—Proceedings of the Conference, Baltimore, MD, USA, 22–27 June 2014; Volume 1.
46. Kieffer, B.; Babaie, M.; Kalra, S.; Tizhoosh, H.R. Convolutional Neural Networks for Histopathology Image Classification: Training vs. Using Pre-Trained Networks. In Proceedings of the 7th International Conference on Image Processing Theory, Tools and Applications, IPTA 2017, Montreal, QC, Canada, 28 November–1 December 2017; Volume 2018.
47. Kumar, J.; Ye, P.; Doermann, D. Structural Similarity for Document Image Classification and Retrieval. *Pattern Recognit. Lett.* **2014**, *43*, 119–126. [CrossRef]
48. Mehmood, Z.; Rashid, M.; Rehman, A.; Saba, T.; Dawood, H.; Dawood, H. Effect of complementary visual words versus complementary features on clustering for effective content-based image search. *J. Intell. Fuzzy Syst.* **2018**, *35*, 5421–5434. [CrossRef]
49. Sicre, R.; Awal, A.M.; Furon, T. Identity Documents Classification as an Image Classification Problem. In *Image Analysis and Processing—ICIAP 2017*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2017; Volume 10485 LNCS.
50. Howard, J.; Ruder, S. Universal Language Model Fine-Tuning for Text Classification. In Proceedings of the ACL 2018—56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), Melbourne, Australia, 15–20 July 2018; Volume 1.
51. Vu, N.T.; Adel, H.; Gupta, P.; Schütze, H. Combining Recurrent and Convolutional Neural Networks for Relation Classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016—Proceedings of the Conference, San Diego, CA, USA, 12–17 June 2016.
52. Adel, H.; Schütze, H. Global Normalization of Convolutional Neural Networks for Joint Entity and Relation Classification. In Proceedings of the EMNLP 2017—Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017.
53. Yin, W.; Schütze, H.; Xiang, B.; Zhou, B. ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs. *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 259–272, Erratum in *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 566–567. [CrossRef]