APPENDIX

A. Experimental Settings

1) Dataset: **PubMed** corpus contains 2.55M abstracts, including 22.9M sentences, related to brain science. 1.21M (47.5%) documents mention at least one pre-defined species using the mention-based annotation. The labels of these documents may not be complete, as the abstract may not mention all species. These documents can be used for further research in knowledge linking and extraction projects. We sample 5,040/778/775 documents as the experimental train/dev/test datasets. Figure 1 (a) visualizes the distribution of sentence number of the abstract. The x and y axes are the sentence number in a scientific work and the count of scientific works that have the corresponding number of sentences respectively. Each document averagely contains 8.9 sentences. Figure 1 (b) visualizes the sentence length distribution. Figure 2 (a) visualizes the species distribution. “Human”, “Mouse” and “Rat” are more frequent labels.

PMC Mention corpus consists of 0.43M articles, including 54.3M sentences, related to brain science. 0.36M (83.5%) documents mention at least one pre-defined species. Annotating the entire corpus is costly and time-consuming, so we sample 1,427/204/195 documents as the train/dev/test datasets for our experiments. Figure 1 (c) visualizes the distribution of sentence number of the paper. The sentence distribution varies over a wide range (14-3087). Long documents occupy a small portion, so we merge the documents with more than 600 sentences. The criteria of this corpus is the species mention. Each document averagely contains 205.6 sentences. Figure 1 (d) visualizes the sentence length distribution. Figure 2 (b) visualizes the species distribution. “Human”, “Mouse”, “Rabbit” and “Rat” are more frequent labels.

PMC Semantics dataset uses the same documents of PMC Mention dataset. We let domain experts annotate these documents. The criteria of this version are based on expert knowledge. Figure 2 (b) visualizes the species distribution. “Human”, “Mouse”, “Not applicable” and “Cell” are more frequent labels.

2) Evaluation: In single-label classification (1-of-n), the prediction can be either correct or wrong. Compared with the single-label classification, MLC is unique since the prediction can be partially correct [1]. MLC requires different evaluation metrics to evaluate the partially correct. Following [2], [3], [4], we adopt the Hamming loss, micro-F1 score. Besides, we also measure the macro-F1 score and F1 per document. F1 per document would also be informative to measure document-level performance. This metric is calculated by averaging the precision, recall and F1 of each document.

$$\text{Hamming} = \frac{1}{|N| \cdot |L|} \sum_{i=1}^{|N|} \sum_{j=1}^{|L|} \text{xor}(y_{(i,j)}, t_{(i,j)}) \qquad (1)$$

A. Hamming Loss calculates the fraction of wrong labels. The lower the hamming loss, the better the performance is, as shown in formula (1). For an ideal classifier, the Hamming loss is 0.

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (2)$$

B. Micro-F1 is the harmonic mean of micro-precision and micro-recall as formula (2). This metric calculates metrics globally by counting the total true positives, false negatives and false positives. This metric aggregates the contributions of all classes.

C. Macro-F1 computes the metric independently for each class and then take the average. This measurement treats all classes equally. We can evaluate the overall model performance for all classes.

3) Implementation Details: Table I reports the main hyperparameters. We train the 200-D GloVe embedding on the whole PubMed and PMC corpora (3M documents). We did not update the pre-trained word embeddings during model training. For the character embeddings, we initialize each character as a 25-D vector. If using character Bi-LSTM, we set 50-D hidden state. If using character CNN, the convolution kernel width is 3, and we use max-pooling to generate 100-D vector representation. The Bi-LSTM dimension of encoder and decoder is 200-D. We use the Adam algorithm [5] to train the model. The initial learning rate is 0.001. The size of species embedding is 200-D. We limit the sentence length to 128 and section length to 512 tokens. We conducted experiments on an Intel(R) Xeon(R) CPU E7-4830 v3 @ 2.10GHz (Mem: 976G) and the GPU Tesla K40c (12G) and TITAN RTX (24G).

TABLE I: The hyperparameter configuration

Hyperparameters	**Value**
character embedding	25
CNN kernel width	3
encoder LSTM	100
decoder LSTM	100
dropout	0.5
word embedding	GloVe.PubMed.200D
epoch	100

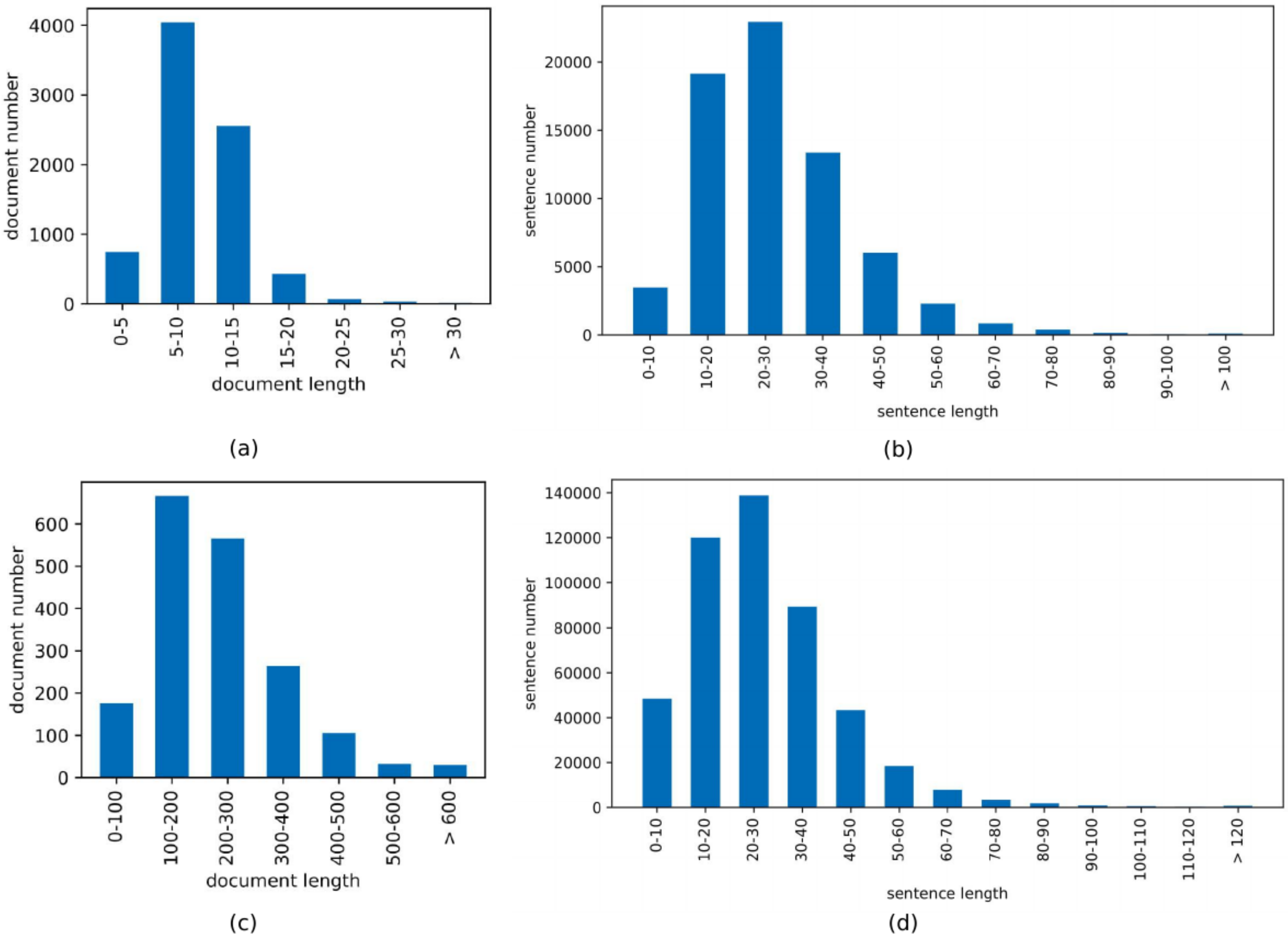


Fig. 1: Dataset visualization, where (a) PubMed sentence distribution of each document (b) PubMed sentence length distribution (c) PMC sentence distribution of each document (d) PMC sentence length distribution

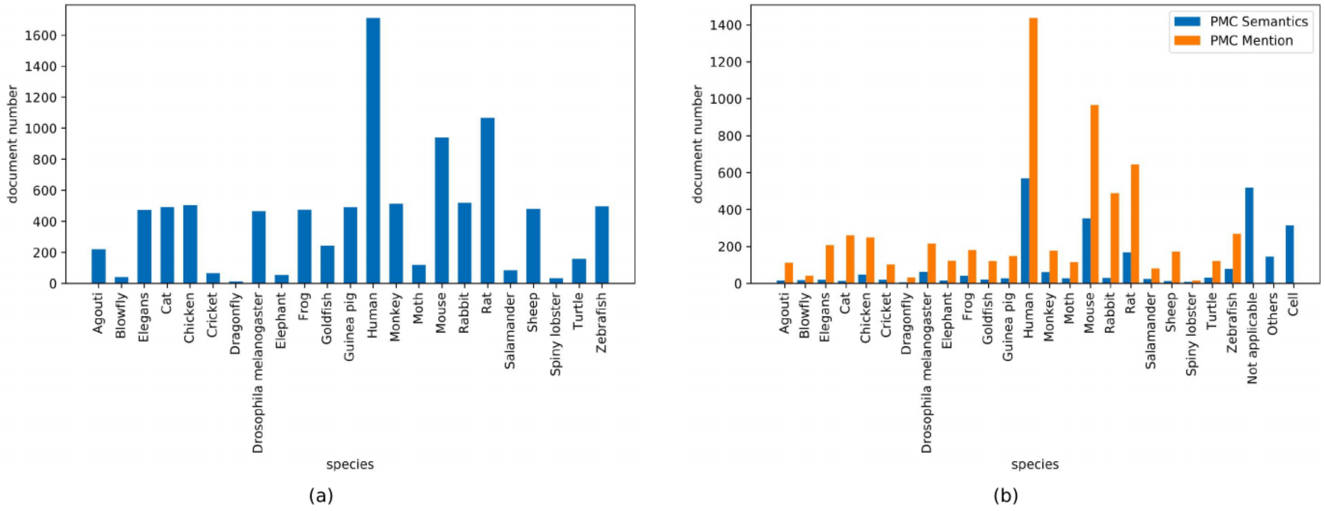
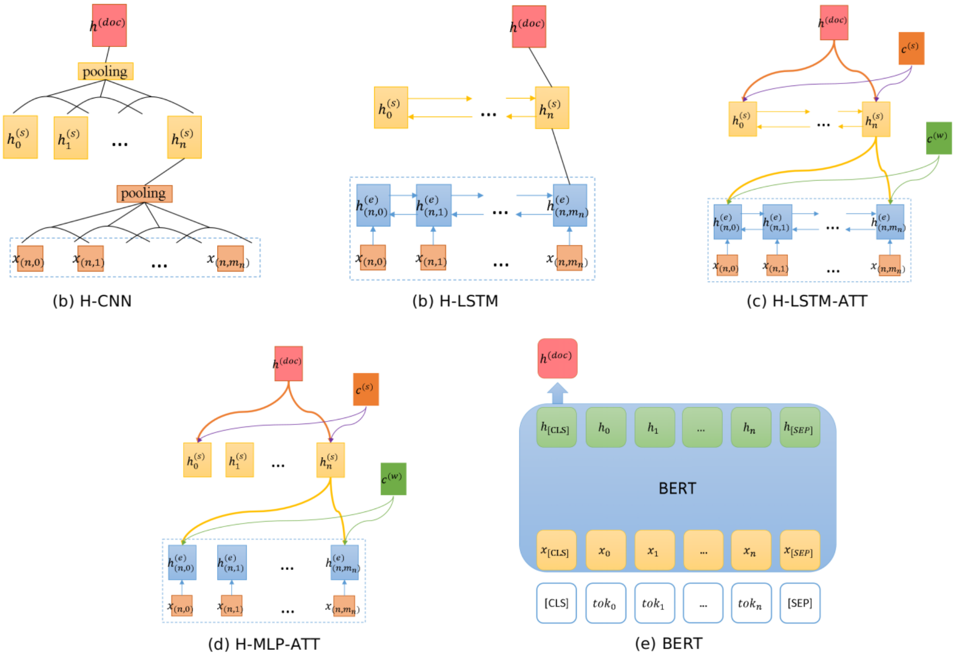


Fig. 2: Species distribution, where (a) species distribution of the PubMed dataset (b) species distribution of the PMC dataset

B. Analysis and Discussion

1) Ablation Study: To analyze the contributions and effects of different components, we perform ablation studies on the PubMed dataset. The performance degrades by 1.83% micro-F1 without **s**entence-level **att**ention (s-att). This is because the model cannot consider the sentence-level structure. The



(b) H-CNN

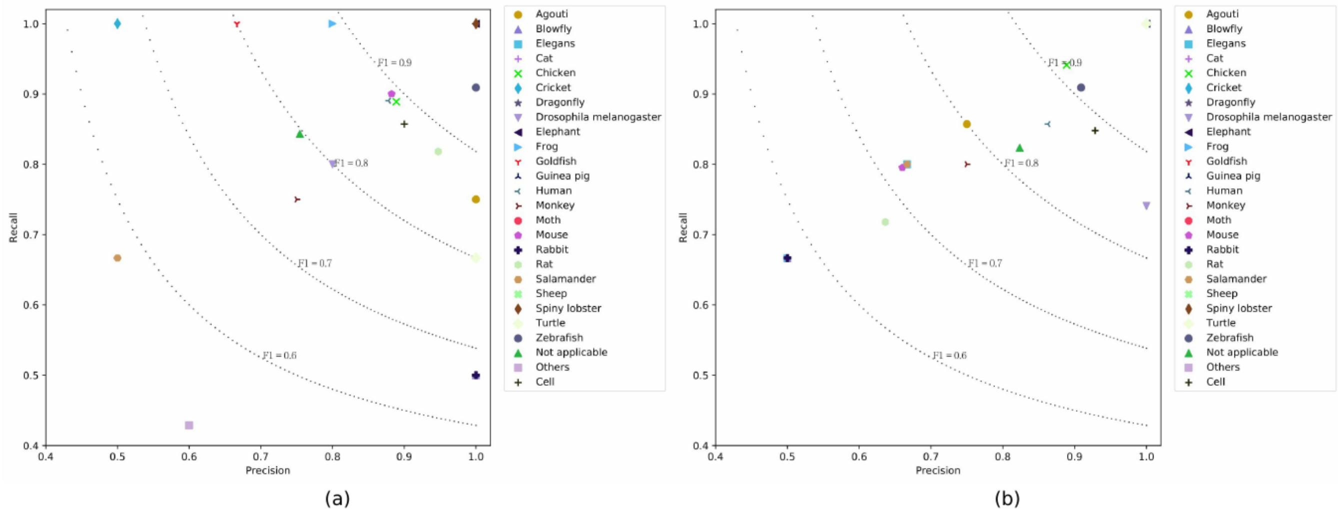
(b) H-LSTM

(c) H-LSTM-ATT

(d) H-MLP-ATT

(e) BERT

Fig. 3: Architectures of baseline models



(a)

(b)

Fig. 4: Prediction results on the PMC Semantics dataset, where (a) prediction results using SeqC + Discourse (b) prediction results using BERT

single-level attention only considers the word sequence, which assumes all sentences of a document are equally relevant for word selection. This setting limits the performance. When we remove the word-level attention (w-att), the performance drops by 2.02% micro-F1 and 4.28% macro-F1. This setting assumes that the contribution of all words in a sentence is the same, but the contribution of different sentences is different.

When we remove the HAD mechanism (s-att and **w**ord-level **att**ention (w-att)), the performance drops by 3.62% micro-F1 and 4.61% macro-F1. This is because the model only uses the document vector to generate species and the decoder cannot attend to the document. When we remove the HAD mechanism and the decoder, the performance drops by 3.71% micro-F1 and 8.33% macro-F1. This is because the model becomes H-LSTM. The memory of a single document vector is limited.

2) Results of Different Species: It is instructive to analyze the prediction result of different species. Figure 2 (a) in article body, Figure 3 (a) in article body and Figure 4 (a) visualize the class-aware prediction results. The x- and y-axes represent the precision and recall respectively. The dotted lines denote the contours of the F1. For the PubMed dataset, we found "Dragonfly", "Blowfly", "Agouti", "Elegans" and "Human" are more easy to predict. The "Spiny lobster", "Rabbit", "Cat" and "Goldfish" are more problematic. For the PMC Mention dataset, we observe the "Human" and "Mouse" are easier to extract. The "Sheep", "Guinea pig", "Cricket" and "Cat" are more problematic. BERT achieves the higher scores than other baselines. For the PMC Semantics dataset, we observe the "Elephant", "Spiny lobster", "Zebrafish" are easier to extract. The "Salamander" and "Others" are more problematic. We observe the prediction results are highly correlated to the class distribution.

As shown in Figure 2 (b), when we let experts annotate the corpus, the class imbalance problem has become more serious. This poses a challenge to the model. This phenomenon often occurs. Different versions of the annotated data have different class distributions. The forecasting of the results of the corpus annotation is important.

TABLE II: The ablation results on the PubMed dataset

Model	Hamming	Micro-F1	Macro-F1
SeqC	0.0247	83.57	82.42
–s-att	0.0274	81.74	82.06
–w-att	0.0274	81.55	78.14
–HAD (s-att,w-att)	0.0300	79.95	77.81
–HAD (s-att,w-att), decoder	0.0292	79.86	74.09

C. Case Study

It is instructive to analyze how the attention mechanism extracts SOIs to predict species. We choose two abstracts [6],[7] to visualize the attention distribution, as shown in Figure 5 and Figure 6. When the model predicts different species, it attends to different parts of the document. We restore the species names in the figure to better understand the samples. These species are marked with underlined stars.

For the first sample, this model first predicts "Human" by using the document representation. We observe this class is not mentioned in the abstract but is mentioned in the text so the "Human" can be assigned to this paper. This means our model can help infer more complete species. Some terms are potential topics in human-related research, e.g., "Huntington's disease", "Cognitive dysfunction", "huntingtin gene", "monogenetic disorder", etc. Figure 5 (a) visualizes the attention distribution when predicting "Human". The attention distribution ("transgenic HD, N171-82Q, HD, neural, WT-NPCs, iPSCs") also contains information about the next species to be predicted, as this decoder sequentially models the correlation between species. When predicting "Mouse", the attention weight of "monogenetic, N171-82Q, neural progenitor, NPCs, pluripotent" increases and the weight of "iPSCs, WT-NPCs" decreases, as shown in Figure 5 (b). When predicting "EOS", token weights are distributed over all emphasized words and are most distracting, as shown in Figure 5 (c). This shows that the model attends to different words when predicting different species. The model also considers the correlation between labels and retains historical memory. However, this model misses "Monkey".

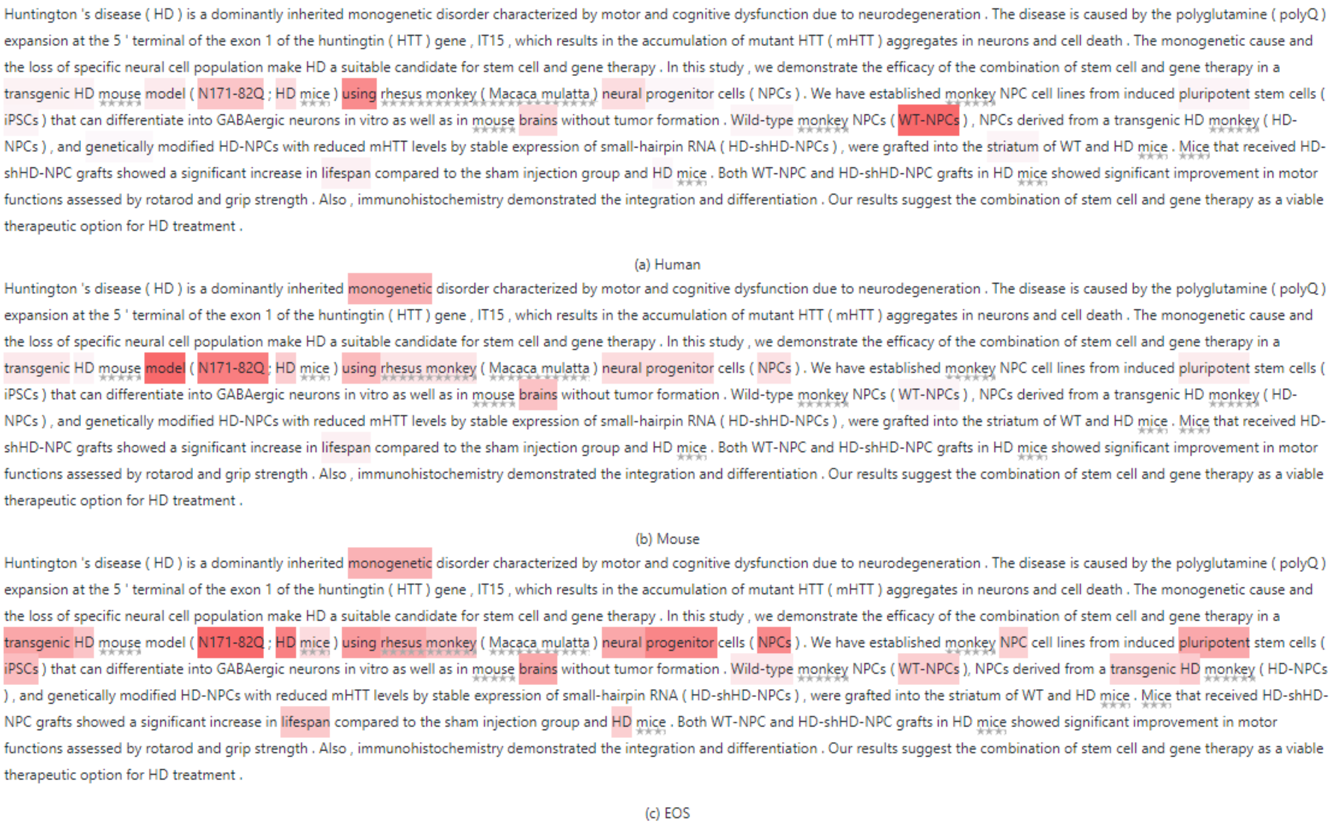
For the second sample, when predicting "Human", the model uses the document representation and attends to "neural, experimentation, nervous system, T-UCRs". When predicting "Monkey", the attention weights of "T-UCRs" and masked species words ("rhesus monkey") are increased. When predicting "Mouse", the weights of "T-UCRs, nervous systems, neural stem" are increased. When predicting "Rat", the weights of "nervous systems, neural stem" are decreased. When predicting "EOS", token weights are most distracting.

Appendix

A. Standard of The Semantic-based Annotation

Most labels denote single species, e.g., rat, mouse, C. elegans and D. Melanogaster, etc. We also need to use different levels of species as the label. For example, moth contains many types of species. A specific moth has been studied in only a few articles, so considering specific moth cannot find valuable related studies. Studies on different moths have many common features in the paper. Treating all the moths as one label can help summarize the research of moth while distinguishing all species will be impossible and will not generate more valuable analysis.

Most classes (species) are organisms, but we add a special class "cell". As long as the article involves experiments on living cells, the article is assigned to "cell". The reason for adding the "cell" class is that most cell-centric experiments use common methodologies, while the organisms that provide these cells are relatively unimportant. As long as the origin of the cell can be found in the article, the document is also assigned to the species of the original organism. Note that sometimes the original organism of cell lines cannot be directly extracted from the article. For example, in some researches, the article merely mentions that the experiment is carried out in HEK293 cell line. If we google HEK293, we could find that it is a human embryo kidney cell line from the human. Although we cannot read it from the article alone, this article will still be assigned to "cell" and "human".



Huntington 's disease (HD) is a dominantly inherited monogenetic disorder characterized by motor and cognitive dysfunction due to neurodegeneration . The disease is caused by the polyglutamine (polyQ) expansion at the 5 ' terminal of the exon 1 of the huntingtin (HTT) gene , IT15 , which results in the accumulation of mutant HTT (mHTT) aggregates in neurons and cell death . The monogenetic cause and the loss of specific neural cell population make HD a suitable candidate for stem cell and gene therapy . In this study , we demonstrate the efficacy of the combination of stem cell and gene therapy in a transgenic HD mouse model (N171-82Q ; HD mice) using rhesus monkey (Macaca mulatta) neural progenitor cells (NPCs) . We have established monkey NPC cell lines from induced pluripotent stem cells (iPSCs) that can differentiate into GABAergic neurons in vitro as well as in mouse brains without tumor formation . Wild-type monkey NPCs (WT-NPCs) , NPCs derived from a transgenic HD monkey (HD-NPCs) , and genetically modified HD-NPCs with reduced mHTT levels by stable expression of small-hairpin RNA (HD-shHD-NPCs) , were grafted into the striatum of WT and HD mice . Mice that received HD-shHD-NPC grafts showed a significant increase in lifespan compared to the sham injection group and HD mice . Both WT-NPC and HD-shHD-NPC grafts in HD mice showed significant improvement in motor functions assessed by rotarod and grip strength . Also , immunohistochemistry demonstrated the integration and differentiation . Our results suggest the combination of stem cell and gene therapy as a viable therapeutic option for HD treatment .

(a) Human

Huntington 's disease (HD) is a dominantly inherited monogenetic disorder characterized by motor and cognitive dysfunction due to neurodegeneration . The disease is caused by the polyglutamine (polyQ) expansion at the 5 ' terminal of the exon 1 of the huntingtin (HTT) gene , IT15 , which results in the accumulation of mutant HTT (mHTT) aggregates in neurons and cell death . The monogenetic cause and the loss of specific neural cell population make HD a suitable candidate for stem cell and gene therapy . In this study , we demonstrate the efficacy of the combination of stem cell and gene therapy in a transgenic HD mouse model (N171-82Q ; HD mice) using rhesus monkey (Macaca mulatta) neural progenitor cells (NPCs) . We have established monkey NPC cell lines from induced pluripotent stem cells (iPSCs) that can differentiate into GABAergic neurons in vitro as well as in mouse brains without tumor formation . Wild-type monkey NPCs (WT-NPCs) , NPCs derived from a transgenic HD monkey (HD-NPCs) , and genetically modified HD-NPCs with reduced mHTT levels by stable expression of small-hairpin RNA (HD-shHD-NPCs) , were grafted into the striatum of WT and HD mice . Mice that received HD-shHD-NPC grafts showed a significant increase in lifespan compared to the sham injection group and HD mice . Both WT-NPC and HD-shHD-NPC grafts in HD mice showed significant improvement in motor functions assessed by rotarod and grip strength . Also , immunohistochemistry demonstrated the integration and differentiation . Our results suggest the combination of stem cell and gene therapy as a viable therapeutic option for HD treatment .

(b) Mouse

Huntington 's disease (HD) is a dominantly inherited monogenetic disorder characterized by motor and cognitive dysfunction due to neurodegeneration . The disease is caused by the polyglutamine (polyQ) expansion at the 5 ' terminal of the exon 1 of the huntingtin (HTT) gene , IT15 , which results in the accumulation of mutant HTT (mHTT) aggregates in neurons and cell death . The monogenetic cause and the loss of specific neural cell population make HD a suitable candidate for stem cell and gene therapy . In this study , we demonstrate the efficacy of the combination of stem cell and gene therapy in a transgenic HD mouse model (N171-82Q ; HD mice) using rhesus monkey (Macaca mulatta) neural progenitor cells (NPCs) . We have established monkey NPC cell lines from induced pluripotent stem cells (iPSCs) that can differentiate into GABAergic neurons in vitro as well as in mouse brains without tumor formation . Wild-type monkey NPCs (WT-NPCs), NPCs derived from a transgenic HD monkey (HD-NPCs) , and genetically modified HD-NPCs with reduced mHTT levels by stable expression of small-hairpin RNA (HD-shHD-NPCs) , were grafted into the striatum of WT and HD mice . Mice that received HD-shHD-NPC grafts showed a significant increase in lifespan compared to the sham injection group and HD mice . Both WT-NPC and HD-shHD-NPC grafts in HD mice showed significant improvement in motor functions assessed by rotarod and grip strength . Also , immunohistochemistry demonstrated the integration and differentiation . Our results suggest the combination of stem cell and gene therapy as a viable therapeutic option for HD treatment .

(c) EOS

Fig. 5: Visualization of SOIs when predicting (a) Human (b) Mouse and (c) EOS where redness indicates attention and the stars below the text indicate the masked species

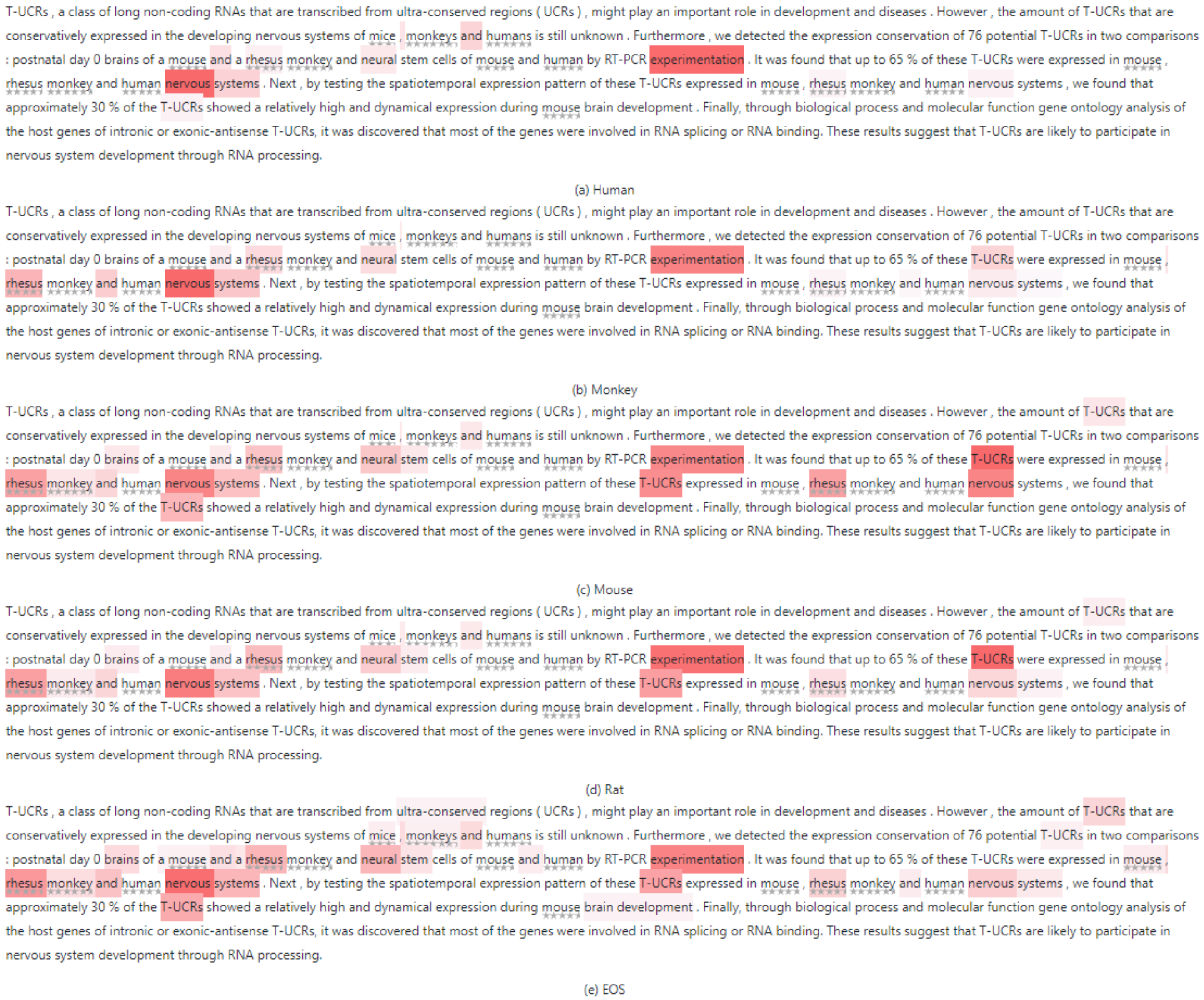
Organisms not included in the pre-defined species are labeled as “others”. If an article cannot be assigned to any of the species including “cell” and “others”, then the article will be labeled as “not applicable”. In this case, it is not appropriate to assert that an article is related to a particular creature. For example, an article develops research method of how to measure the glucose concentration of a solution, or an article generates a phylogenetic tree of all organisms in an order, or an article reports all species found in a certain ecosystem.

We determine whether an organism is associated with an article based on the following criteria:

(1) If an in vivo experiment is carried out on an organism, the article is assigned to the organism. If a living organism is used in the experiment as a tool, the article is not assigned to the organism. For example, in [8], salamander is used as stimulation to induce an effect on the frog. This article will be only assigned to “frog” and not to “salamander”.

(2) Review articles are not assigned to any organism. First, the description of organisms in reviews also exists in research articles, while they do not indicate that the research articles are based on the organism. For example, if a research article is investigating the function of a gene in Drosophila Melanogaster, it would probably mention the research of the gene in mammals [9]. However, this article is not based on mammals. Reviews would make a similar description about the gene. If an article is considered to be mammal-based, then the same description will indicate that the article of investigating Drosophila gene is based on the mammal. This is unreasonable. Some reviews discuss some organisms extensively [10], and it seems reasonable to say that the review is based on a certain organism. However, the distinction between reviews that discuss certain organisms [11] and reviews that discuss a phenomenon [12] related to multiple organisms can be very vague.

Second, sometimes, in a review paper, it is impossible to tell which kind of organism the description is based on. Reviews sometimes do not mention which organism is related to the research. For example, in this review [13], many descriptions do not contain specific organism names. However, in research articles, researches based on certain organisms have specific methodologies and keywords. Analyzing the review paper may not identify the organism. However, analyzing the articles cited in the review paper can easily tell us which organism is related to the review, so excluding review from our research would not diminish the significance.



T-UCRs , a class of long non-coding RNAs that are transcribed from ultra-conserved regions (UCRs) , might play an important role in development and diseases . However , the amount of T-UCRs that are conservatively expressed in the developing nervous systems of mice , monkeys and humans is still unknown . Furthermore , we detected the expression conservation of 76 potential T-UCRs in two comparisons : postnatal day 0 brains of a mouse and a rhesus monkey and neural stem cells of mouse and human by RT-PCR experimentation . It was found that up to 65 % of these T-UCRs were expressed in mouse , rhesus monkey and human nervous systems . Next , by testing the spatiotemporal expression pattern of these T-UCRs expressed in mouse , rhesus monkey and human nervous systems , we found that approximately 30 % of the T-UCRs showed a relatively high and dynamical expression during mouse brain development . Finally, through biological process and molecular function gene ontology analysis of the host genes of intronic or exonic-antisense T-UCRs, it was discovered that most of the genes were involved in RNA splicing or RNA binding. These results suggest that T-UCRs are likely to participate in nervous system development through RNA processing.

(a) Human

T-UCRs , a class of long non-coding RNAs that are transcribed from ultra-conserved regions (UCRs) , might play an important role in development and diseases . However , the amount of T-UCRs that are conservatively expressed in the developing nervous systems of mice , monkeys and humans is still unknown . Furthermore , we detected the expression conservation of 76 potential T-UCRs in two comparisons : postnatal day 0 brains of a mouse and a rhesus monkey and neural stem cells of mouse and human by RT-PCR experimentation . It was found that up to 65 % of these T-UCRs were expressed in mouse , rhesus monkey and human nervous systems . Next , by testing the spatiotemporal expression pattern of these T-UCRs expressed in mouse , rhesus monkey and human nervous systems , we found that approximately 30 % of the T-UCRs showed a relatively high and dynamical expression during mouse brain development . Finally, through biological process and molecular function gene ontology analysis of the host genes of intronic or exonic-antisense T-UCRs, it was discovered that most of the genes were involved in RNA splicing or RNA binding. These results suggest that T-UCRs are likely to participate in nervous system development through RNA processing.

(b) Monkey

T-UCRs , a class of long non-coding RNAs that are transcribed from ultra-conserved regions (UCRs) , might play an important role in development and diseases . However , the amount of T-UCRs that are conservatively expressed in the developing nervous systems of mice , monkeys and humans is still unknown . Furthermore , we detected the expression conservation of 76 potential T-UCRs in two comparisons : postnatal day 0 brains of a mouse and a rhesus monkey and neural stem cells of mouse and human by RT-PCR experimentation . It was found that up to 65 % of these T-UCRs were expressed in mouse , rhesus monkey and human nervous systems . Next , by testing the spatiotemporal expression pattern of these T-UCRs expressed in mouse , rhesus monkey and human nervous systems , we found that approximately 30 % of the T-UCRs showed a relatively high and dynamical expression during mouse brain development . Finally, through biological process and molecular function gene ontology analysis of the host genes of intronic or exonic-antisense T-UCRs, it was discovered that most of the genes were involved in RNA splicing or RNA binding. These results suggest that T-UCRs are likely to participate in nervous system development through RNA processing.

(c) Mouse

T-UCRs , a class of long non-coding RNAs that are transcribed from ultra-conserved regions (UCRs) , might play an important role in development and diseases . However , the amount of T-UCRs that are conservatively expressed in the developing nervous systems of mice , monkeys and humans is still unknown . Furthermore , we detected the expression conservation of 76 potential T-UCRs in two comparisons : postnatal day 0 brains of a mouse and a rhesus monkey and neural stem cells of mouse and human by RT-PCR experimentation . It was found that up to 65 % of these T-UCRs were expressed in mouse , rhesus monkey and human nervous systems . Next , by testing the spatiotemporal expression pattern of these T-UCRs expressed in mouse , rhesus monkey and human nervous systems , we found that approximately 30 % of the T-UCRs showed a relatively high and dynamical expression during mouse brain development . Finally, through biological process and molecular function gene ontology analysis of the host genes of intronic or exonic-antisense T-UCRs, it was discovered that most of the genes were involved in RNA splicing or RNA binding. These results suggest that T-UCRs are likely to participate in nervous system development through RNA processing.

(d) Rat

T-UCRs , a class of long non-coding RNAs that are transcribed from ultra-conserved regions (UCRs) , might play an important role in development and diseases . However , the amount of T-UCRs that are conservatively expressed in the developing nervous systems of mice , monkeys and humans is still unknown . Furthermore , we detected the expression conservation of 76 potential T-UCRs in two comparisons : postnatal day 0 brains of a mouse and a rhesus monkey and neural stem cells of mouse and human by RT-PCR experimentation . It was found that up to 65 % of these T-UCRs were expressed in mouse , rhesus monkey and human nervous systems . Next , by testing the spatiotemporal expression pattern of these T-UCRs expressed in mouse , rhesus monkey and human nervous systems , we found that approximately 30 % of the T-UCRs showed a relatively high and dynamical expression during mouse brain development . Finally, through biological process and molecular function gene ontology analysis of the host genes of intronic or exonic-antisense T-UCRs, it was discovered that most of the genes were involved in RNA splicing or RNA binding. These results suggest that T-UCRs are likely to participate in nervous system development through RNA processing.

(e) EOS

Fig. 6: Visualization of SOIs when predicting (a) Human (b) Monkey (c) Mouse (d) Rat and (e) EOS where redness indicates attention and the stars below the text indicate the masked species

(3) If experiments in the article are based on cells, body fluid, or other body parts related to a certain organism, then the article is considered to be based on the organism. If a part of the organism is used as a tool rather than a research target, this article is not considered to be based on the organism, e.g., using antibody from rabbit, sheep, cattle, rat and hamster, or using serum from cattle, or using cornmeal and wheat in food.

(4) If the protein is used in the research and the protein is indeed the research target, then the research is considered to be based on the species of the protein. If the protein is expressed in another organism or cell, then the paper is usually considered to be based on the target organism or the organism of the cell. Because in this case, the protein has complicated interaction with the organism or cell. For most researches on this topic, observing the effect of the expression is a very important part of the research. Thus, the expression organism or cell cannot be merely considered as a tool. But if the expression organism or cell is just used to produce the protein for further experiments, like crystallization or enzyme catalysis, then they are not considered to be related to the article. If the experiment involves small molecule products like sugar, cellulose or other substrates that are not specifically produced by a species, then the article is not assigned to the organism.

(5) For researches on experimental methods, as long as the articles involve a demonstration or test of experimental methods on organisms, the study is considered to be based on the organism.

(6) For bioinformatics or systematic research, if the research focuses on a few organisms, e.g., comparing features between mouse, rat and human, and discussing them intensively, then

the article is labeled with all these organisms. If the research draws conclusion on certain organism, then the article is labeled with the organism. If the research involves a lot of organisms and does not focus on a few of them, then the article is not labeled with the organism. For example, a research finds a lot of species through systematic searching a screening or collects all species in an orderly manner and forms a phylogenetic tree.

(7) Whether an article can be assigned to an organism still has gray area. For example, a research evaluates a protein produced by different kinds of cells belonging to different organisms, or a research uses DNA sequences from different organisms to deduce the function of a gene. Although the specific range of organism in these researches is uncertain, the academic tradition and methodology of these researches are explicit. In these cases, we insist that organisms need to best describe the academic tradition and methodology of the researches. We will use the same principles to label articles that are similar in academic traditions and methodologies.

B. Related Work

BioNLP has achieved substantial progress on many tasks [14], [15], [16], such as named entity recognition, information extraction, information retrieval, corpora annotation, evaluation, etc. These researches open up opportunity to integrate biomedical text mining with knowledge engineering and data mining. There are some researches on text mining in the genomics domain [17], e.g., identifying gene/protein names and their relations. [18] introduce the methods and challenges in many aspects of health and biomedical information retrieval systems. [19] describe the role of biomedical ontologies in knowledge management, data integration and decision support. There are some ontologies, such as SNOMED CT, the Logical Observation Identifiers, Names, and Codes (LOINC), the Foundational Model of Anatomy, the Gene Ontology, RxNorm, the National Cancer Institute Thesaurus, the International Classification of Diseases, the Medical Subject Headings (MeSH) and the Unified Medical Language System (UMLS). [20] introduce the shared principles governing ontology development in the Open Biomedical Ontologies (OBO). [21], [22], [23] use microarray technology and Gene Ontology (GO) terms to analyze the gene expression to characterize biological processes and identify the mechanisms that underlie diseases.

[24] proposes the one layer CNN architecture with multiple filter width to encode both task-specific and static vectors. [25] propose a neural network using cross-entropy loss instead of the ranking loss. [26] utilize word embeddings based on CNN to capture label correlations. [27] present a variant of CNN based approach to extreme multi-label text classification. [4] propose a method to ensemble the CNN networks to capture diverse information on different nets.

References

[1] R. Venkatesan and M. J. Er, “Multi-label classification method based on extreme learning machines,” in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, 2014, pp. 619–624.

[2] P. Yang, X. Sun, W. Li, S. Ma, W. Wu, and H. Wang, “Sgm: sequence generation model for multi-label classification,” in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 3915–3926.

[3] M.-L. Zhang and Z.-H. Zhou, “Ml-knn: A lazy learning approach to multi-label learning,” *Pattern recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[4] G. Chen, D. Ye, Z. Xing, J. Chen, and E. Cambria, “Ensemble application of convolutional and recurrent neural networks for multi-label text categorization,” in *Neural Networks (IJCNN), 2017 International Joint Conference on*. IEEE, 2017, pp. 2377–2383.

[5] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[6] I. K. Cho, C. E. Hunter, S. Ye, A. L. Pongos, and A. W. S. Chan, “Combination of stem cell and gene therapy ameliorates symptoms in huntington’s disease mice,” *npj Regenerative Medicine*, vol. 4, no. 1, p. 7, 2019.

[7] J. Zhou, R. Wang, J. Zhang, L. Zhu, W. Liu, S. Lu, P. Chen, H. Li, B. Yin, J. Yuan *et al.*, “Conserved expression of ultra-conserved noncoding rna in mammalian nervous system,” *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1860, no. 12, pp. 1159–1168, 2017.

[8] T. Mori, Y. Kitani, J. Ogihara, M. Sugiyama, G. Yamamoto, O. Kishida, and K. Nishimura, “Histological and ms spectrometric analyses of the modified tissue of bulgy form tadpoles induced by salamander predation,” *Biology open*, vol. 1, no. 4, pp. 308–317, 2012.

[9] C. S. Erdogan, B. W. Hansen, and O. Vang, “Are invertebrates relevant models in ageing research? focus on the effects of rapamycin on tor,” *Mechanisms of ageing and development*, vol. 153, pp. 22–29, 2016.

[10] U. Banerjee, J. R. Girard, L. M. Goins, and C. M. Spratford, “Drosophila as a genetic model for hematopoiesis,” *Genetics*, vol. 211, no. 2, pp. 367–417, 2019.

[11] R. L. Bell, H. J. Sable, G. Colombo, P. Hyytia, Z. A. Rodd, and L. Lumeng, “Animal models for medications development targeting alcohol abuse using selectively bred rat lines: neurobiological and pharmacological validity,” *Pharmacology Biochemistry and Behavior*, vol. 103, no. 1, pp. 119–155, 2012.

[12] R. L. Bell, S. Hauser, Z. A. Rodd, T. Liang, Y. Sari, J. McClintick, S. Rahman, and E. A. Engleman, “A genetic animal model of alcoholism for screening medications to treat addiction,” in *International review of neurobiology*. Elsevier, 2016, vol. 126, pp. 179–261.

[13] I. P. Sudhakaran and M. Ramaswami, “Long-term memory consolidation: The role of rna-binding proteins with prion-like domains,” *RNA biology*, vol. 14, no. 5, pp. 568–586, 2017.

[14] S. Ananiadou and J. McNaught, *Text mining for biology and biomedicine*. Citeseer, 2006.

[15] L. Hunter and K. B. Cohen, “Biomedical language processing: what’s beyond pubmed?” *Molecular cell*, vol. 21, no. 5, pp. 589–594, 2006.

[16] L. J. Jensen, J. Saric, and P. Bork, “Literature mining for the biologist: from information retrieval to biological discovery,” *Nature reviews genetics*, vol. 7, no. 2, p. 119, 2006.

[17] P. Zweigenbaum, D. Demner-Fushman, H. Yu, and K. B. Cohen, “Frontiers of biomedical text mining: current progress,” *Briefings in bioinformatics*, vol. 8, no. 5, pp. 358–375, 2007.

[18] W. Hersh, *Information retrieval: a health and biomedical perspective*. Springer Science & Business Media, 2008.

[19] O. Bodenreider, “Biomedical ontologies in action: role in knowledge management, data integration and decision support,” *Yearbook of medical informatics*, vol. 17, no. 01, pp. 67–79, 2008.

[20] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall *et al.*, “The obo foundry: coordinated evolution of ontologies to support biomedical data integration,” *Nature biotechnology*, vol. 25, no. 11, p. 1251, 2007.

[21] R. K. Curtis, M. Orešič, and A. Vidal-Puig, “Pathways to the analysis of microarray data,” *TRENDS in Biotechnology*, vol. 23, no. 8, pp. 429–435, 2005.

[22] P. Khatri and S. Drăghici, “Ontological analysis of gene expression data: current tools, limitations, and open problems,” *Bioinformatics*, vol. 21, no. 18, pp. 3587–3595, 2005.

[23] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists,” *Nucleic acids research*, vol. 37, no. 1, pp. 1–13, 2008.

[24] Y. Kim, “Convolutional neural networks for sentence classification,” *arXiv preprint arXiv:1408.5882*, 2014.

[25] J. Nam, J. Kim, E. L. Mencía, I. Gurevych, and J. Fürnkranz, “Large-scale multi-label text classification—revisiting neural networks,” in *Joint european conference on machine learning and knowledge discovery in databases*. Springer, 2014, pp. 437–452.

[26] G. Kurata, B. Xiang, and B. Zhou, “Improved neural network-based multi-label classification with better initialization leveraging label co-occurrence,” in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 521–526.

[27] J. Liu, W.-C. Chang, Y. Wu, and Y. Yang, “Deep learning for extreme multi-label text classification,” in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2017, pp. 115–124.