

Recognizing Textual Inference in Mongolian Bar Exam Questions

Garmaabazar Khaltarkhuu ^{1,*}, Biligsai Khan Batjargal ² and Akira Maeda ³

¹ Graduate School of Information Science and Engineering, Ritsumeikan University, Kusatsu 525-8577, Shiga, Japan

² Research Organization of Science and Technology, Ritsumeikan University, Kusatsu 525-8577, Shiga, Japan; biligee@fc.ritsumei.ac.jp

³ College of Information Science and Engineering, Ritsumeikan University, Kusatsu 525-8577, Shiga, Japan; amaeda@is.ritsumei.ac.jp

* Correspondence: garmaabazar@gmail.com

Abstract: This paper examines how to apply deep learning techniques to Mongolian bar exam questions. Several approaches that utilize eight different fine-tuned transformer models were demonstrated for recognizing textual inference in Mongolian bar exam questions. Among eight different models, the fine-tuned bert-base-multilingual-cased obtained the best accuracy of 0.7619. The fine-tuned bert-base-multilingual-cased was capable of recognizing “contradiction”, with a recall of 0.7857 and an F1 score of 0.7674; it recognized “entailment” with a precision of 0.7750, a recall of 0.7381, and an F1 score of 0.7561. Moreover, the fine-tuned bert-large-mongolian-uncased showed balanced performance in recognizing textual inference in Mongolian bar exam questions, thus achieving a precision of 0.7561, a recall of 0.7381, and an F1 score of 0.7470 for recognizing “contradiction”.

Keywords: natural language inference; deep learning; Mongolian bar exam questions; legal analysis



Citation: Khaltarkhuu, G.; Batjargal, B.; Maeda, A. Recognizing Textual Inference in Mongolian Bar Exam Questions. *Appl. Sci.* **2024**, *14*, 1073. <https://doi.org/10.3390/app14031073>

Academic Editor: Andrea Prati

Received: 19 December 2023

Revised: 20 January 2024

Accepted: 24 January 2024

Published: 26 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the World Bank’s Worldwide Governance Indicators, Mongolia has a relatively low quality of governance. Its government effectiveness score is 34.91%, regulatory quality is 42.45%, and rule of law is 45.75% [1]. Additionally, during a speech at the intermediate evaluation discussion panel of the “New Development Mid-term Action Plan”, the Chief Cabinet Secretary of the Mongolian government stated that “Over the past 20 years in Mongolia, the government, ministries, and their agencies have issued 517 short or long-term development plans and strategy papers. “Currently, 203 of these documents are effective, though many of them overlap or contradict each other significantly. Only 132 of them are enforced, which has led to less than 26% efficiency [2]”. Using artificial intelligence (AI) in the analysis of Mongolian government documents is critical, as the current situation produces significant contradictions and overlaps in Mongolian government documents.

On the other hand, at the organizational level, analyzing contracts or legal documents and making decisions at the management level is an important task. The demand is increasing as researchers, lawyers, executives, and managers examine more legal documents with faster and more accurate results. Furthermore, expert decision-making computer systems are devices in demand by professionals and managers to avoid problems and disputes. It is vital to make professional and accurate decisions backed by achievements in AI. Thus, our goal is to construct a decision support system (DSS) to help managers solve complex problems in Mongolia. Such a system is hardly implemented in Mongolia. As far as we know, a deep learning method has not yet been matured for Mongolian legal documents.

Moreover, a comprehensive analysis should be undertaken to deal with Mongolian legal documents instead of deploying a language model as it exists in English. In 1992,

Mongolia adopted civil law, thus discarding the socialist legal system, which had been in place for the past sixty-eight years. However, English-speaking countries, including Canada and the United States of America, follow case law. Depending on the legal system, jurisdiction procedures vary. Therefore, language models developed for the case law cannot be immediately deployed to the Mongolian legal domain, which is based on civil law.

1.1. Motivations of This Research

The increasing demands in Mongolia and the current situation mentioned in Section 1 have prompted us to undertake extensive research to develop deep learning methods to analyze Mongolian legal documents. However, as described in Section 2, the existing research in Mongolian natural language processing (NLP) has left behind the cutting-edge trend and has notable gaps in existing knowledge. Although there is a lack of decent NLP tools for the Mongolian language, analyzing Mongolian legal documents using the rapidly emerging deep learning methods is a pioneering opportunity to become a starting point for the development of advanced systems in the Mongolian legal domain. Therefore, this research is vital to solving real-world problems in Mongolia, and it has a potential impact on the Mongolian legal domain.

1.2. Scope of the Present Paper

This paper focuses on modern Mongolian legal documents written in the Cyrillic script. It does not cover legal documents from the Inner Mongolia Autonomous Region written in traditional Mongolian script.

The Mongolian language is spoken by people in Mongolia, as well as by ethnic Mongols living in China and Russia. Throughout history, Mongols have created and used various writing scripts such as traditional Mongolian, Phags-pa, Horizontal square, Soyombo, Todo, Latin script, and even phonetic writing with Chinese characters [3]. In 1946, a language reform took place, and the Cyrillic script was adopted as the official script for the Mongolian language. This adaptation included two additional characters.

The spelling of modern Mongolian in Cyrillic script was based on the pronunciation of the Khalkha dialect of the largest Mongolian ethnic group [4]. This change was significant because the traditional script preserved the old Mongolian language, while the modern Mongolian in Cyrillic script reflected the pronunciations in modern dialects. Although the spoken language changed as the Mongolian language evolved, the spelling remained unchanged in the traditional Mongolian script. As a result, there are notable differences between the traditional script and the Cyrillic script documents. The Mongolian language is agglutinative, meaning that inflectional suffixes such as plural, case, reflexive, voice, tense, aspect, and mood suffixes are concatenated with the stem. Therefore, stemming is necessary for Cyrillic script. It is important to note that modern Mongolian in Cyrillic script is case-sensitive, while the traditional script is not. Such distinctions affect the use of NLP tools and language resources, as they are not the same for modern and traditional Mongolian.

1.3. Contributions of the Present Paper

To achieve the research goal, this paper examines deep learning techniques for Mongolian legal documents. Particularly, this paper discusses the authors' achievements in recognizing textual inference in Mongolian bar exam questions. We believe that recognizing textual inference is one of the important tasks in our system.

The contributions of this paper can be summarized as follows:

- A creation of a textual inference dataset from Mongolian bar exam questions;
- A pioneering trail to demonstrate fine-tuned transformer models for recognizing textual inference in Mongolian bar exam questions;
- The development of a demo system that can be used to recognize textual inference in Mongolian legal documents by utilizing the above contributions.

Section 2 introduces related work. Existing deep learning tasks and language models for the Mongolian language are also briefed in Section 2. However, research related

to (1) the NLP of texts in the traditional Mongolian script and (2) the analysis of legal documents written in the traditional Mongolian script are not included there. Recognizing textual inference tasks for the Mongolian language are then explained in Section 3. The experimental results are explained in Section 4. Finally, concluding remarks are given in Section 5.

2. Related Work

Deep learning has proven to deliver high performance in a variety of fields [5] and has already achieved near-human performance in various NLP tasks in English [6,7]. In recent years, pretrained language models such as BERT [8] and text-to-text transfer transformer (T5) [9] have provided notable achievements on large-scale general natural language inference (NLI) datasets such as the Stanford Natural Language Inference corpus [10], and the Multi-Genre Natural Language Inference corpus [11]. However, existing Mongolian legal datasets are scarce, so the performance drops when directly adapting the existing pretrained models. OpenAI introduced ChatGPT models, which communicate conversationally to answer questions [12]. However, as far as we know, no research has been conducted to evaluate the performance of ChatGPT in Mongolian language. In general, deep learning approaches to the Mongolian language lag behind due to a lack of research in Mongolian NLP. AI implementation in the Mongolian legal domain is very limited. The following sections summarize the relevant work in recognizing textual inference in the legal domain and Mongolian NLP.

2.1. Recognizing Textual Inference in the Legal Domain

Under the umbrella of the NLI field, various approaches in recognizing textual inference have been applied to the legal domain. Recognizing textual inference in the legal domain is a task of predicting entailment between a given premise, i.e., a law or an article from a law, and the given hypothesis, which is a statement of a legal question.

At the Legal Textual Entailment task of the Competition on Legal Information Extraction/Entailment (COLIEE) [13], various methods, including an ensemble of several BERT fine-tuning runs [14–16], ensemble with data augmentation [14–16], positive and negative sampling [14,15], an ensemble of predicate–argument structures, and a rule-based method [16] were demonstrated for recognizing inference between a legal question and Japanese civil law articles. Bui et al. [17] utilized Fine-tuned LAnguage Net (FLAN) large language models (LLMs) such as FLAN-T5 [18], FLAN-UL2 (unifying language learning) [19], and FLAN-Alpaca-XXL [20] along with data augmentation. They manually selected 56 prompts from the PromptSource library and used them in zero-shot prompting on the above LLMs for predicting the inference of given Japanese civil law problem–article pairs. Onaga et al. [21] extended their previous model [16] and integrated a RoBERTa-based model LUKE-Japanese [22] as a Japanese named entity recognition model (NER) for dealing with anonymized personal names. The COLIEE-2023 training data contain 996 pairs of a legal question and Japanese civil law articles. The performance of each submitted method was evaluated on the test data that contained 101 pairs of a legal question from the most recent Japanese bar exam. However, these methods using the COLIEE use Japanese language-specific approaches, which cannot be directly applied to recognizing textual inference in the Mongolian language.

Moreover, some methods have been proposed to identify a paragraph(s) from existing cases in Canadian case law, which entails a given new case. Nguyen et al. [23] utilized the sequence-to-sequence model MonoT5 [24] and fine-tuned it with hard negation mining and ensemble methods, which search hyperparameters to find the optimal weight for each checkpoint. However, such models developed for the case law system cannot be immediately applied to recognizing textual inference in Mongolian bar exam questions due to jurisdictional differences.

2.2. Mongolian NLP

This section summarizes research on NLP in the Mongolian language. Although deep learning has become attractive in recent years, little research in deep learning has been conducted on the Mongolian language. Ariunaa and Munkhjargal utilized a recurrent neural network for the sentiment analysis of Mongolian tweets [25]. Battumur et al. trained BERT to correct Mongolian spelling errors [26]. Dashdorj et al. utilized BiLSTM and a convolutional neural network to classify public complaints addressed to government agencies [27].

Following is a brief explanation of traditional NLP approaches for the Mongolian language, Choi and Tsend determined the appropriate size of the Mongolian general corpus using the Heaps' law and type token ratio. They concluded that an appropriate size for a Mongolian general corpus is 39–42 million tokens [28]. To tag Mongolian parts of speech (POS), Lkhagvasuren et al. utilized a neural network model with a multilayer perceptron [29]. Jaimai and Chimeddorj utilized a hidden Markov model with a bigram [30]. Ivanov et al. utilized a minimum edit distance and a word n-gram with a back-off to identify Mongolian spelling errors and suggest alternative correct words [31]. Batsuren et al. built a Mongolian WordNet [32]. To determine Mongolian word sense disambiguation, Bataa and Altangerel utilized a "one sense per collocation" algorithm [33]. Jaimai et al. built a Mongolian morphological analyzer [34] utilizing the program PC-KIMMO. Dulamragchaa et al. [35] built Mongolian phonological rules and lexicon for a PC-KIMMO program. Chagnaa and Adiyatseren improved the Mongolian two-level rules for PC-KIMMO [36]. Munkhjargal et al. built a finite state Mongolian morphological transducer [37]. Enkhbayar et al. developed a stemming method for Mongolian nouns and verbs [38]. Khaltar and Fujii developed a Mongolian lemmatization method [39]. Nyandag et al. tested a keyword extraction method utilizing cosine similarity and TF-IDF [40]. Munkhjargal et al. created a Mongolian NER system [41]. Khaltar et al. [42,43] created a method for extracting loan words from the Mongolian corpora. Lkhagvasuren and Rentsendorj built an open information extraction system that extracts relation tuples from Mongolian text [44]. In their research, Damiran and Altangerel tested (1) a decision tree algorithm [45] to classify Mongolian novels according to their authors and (2) a naive Bayesian classifier to classify Mongolian news articles [46]. Ehara et al. developed a transfer-based Mongolian-to-Japanese machine translation system [47] using the ChaSen, which is a Japanese morphological analyzer. Enkhbayar et al. examined the ambiguity degree in the Japanese-to-Mongolian translation of functional expressions [48].

Mongolian NLP research is limited to those mentioned above, and none of them has considered analyses of Mongolian legal documents. We believe that analyzing the Mongolian bar exam questions is a good opportunity to test deep learning methods and become a starting point for the development of advanced systems in the Mongolian legal domain.

2.3. Existing Deep Learning-Based Tasks and Language Models for the Mongolian Language

Since BERT has been shown to perform well, Erdene-Ochir revealed Mongolian BERT models [49] trained on approximately 500 million words extracted from Mongolian media and a Mongolian Wikipedia dump. Yadamsuren released Mongolian RoBERTa base [50], Mongolian RoBERTa large [51], Mongolian ALBERT [52], and Mongolian GPT2 [53], which were trained on the same dataset as the Mongolian BERT models.

Gunchinish [54] revealed text classification tasks that classify Mongolian news articles by fine-tuning bert-base-mongolian-cased [55] and bert-large-mongolian-uncased [56] models. Bataa [57] finetuned the bert-base-mongolian-cased [55] model for the Mongolian NER task. Conneau et al. trained a transformer-based masked language model on one hundred languages and included Mongolian CommonCrawl data [58]. However, they did not disclose the results for the Mongolian language. Google publicized an un-normalized multilingual model named BERT-base-multilingual-cased, which does not perform any normalization on the input and additionally included the Mongolian language [59].

There are not many NLP tasks available in the Mongolian language that have utilized masked language models such as BERT. Thus, this paper aims to demonstrate an NLI task for recognizing textual inference in Mongolian bar exam questions.

2.4. Existing Datasets of Mongolian Documents

This section summarizes the existing datasets and corpora available in the Mongolian language. As of 11 January 2024, the following Mongolian datasets are available at the public domain: the Mongolian Wordnet of 26,875 words, 2979 glosses, 23,665 synsets, and 213 examples [60]; UniMorph's derivational morphology paradigms of 1410 lemmas, 1629 derivations, and 229 derivational suffixes [61]; MorphyNet's inflectional morphology paradigms of 2085 lemmas, 14,592 inflected forms, and 35 morpheme segmentation [62]; Mongolian 75K news in nine categories [63]; raw text of Mongolian news that contains approximately 700 million words [64]; the 397M Common Crawl Mongolian dataset [65]; a dataset of 80,036 public complaints to government agencies of five categories: (1) feedback, (2) complaint, (3) criticism, (4) appreciation, and (5) claim [66]; 10,162 manually annotated sentences from Mongolian politics and sport news with four tags: LOCATION, PERSON, ORGANIZATION, and MISC [67]; 223,171 personal names [68]; 90,333 clan or family names [69]; 192,771 company names [70]; 353 names of Mongolian provinces, districts, and villages [71]; 195 country names and their capital names in Mongolian [72]; and 527 Mongolian abbreviations [73].

Mongolian Legal Documents

Because our target is the Mongolian legal domain, this section summarizes the existing Mongolian legal documents. Among the Mongolian 75K news articles described in Section 2.4, 8285 news articles are in the category "legal", which are 10.9% of the news articles. In addition, many Mongolian legal documents have been made publicly available in human-readable digital formats such as HTML, PDF, or Microsoft Word documents. The legal documents include (1) the Constitution of Mongolia; (2) government resolutions; (3) resolutions of the Mongolian Parliament; (4) resolutions of the self-governing bodies; (5) Mongolian laws; (6) decrees of the ministries; (7) international treaties and contracts endorsed by Mongolia; (8) government agencies' regulations; (9) decisions of the Constitutional Court; (10) President's decrees; (11) Supreme Court resolutions; (12) various committees, councils, and collectives' decisions; (13) decisions by the heads of the organizations appointed by the Parliament; (14) the mayor of Ulaanbaatar's and provinces governors' orders; and (15) the General Judge Council's resolutions, which can be found on the public domain website "Unified Legal Information System" [74] of the National Legal Institute of Mongolia. As of 12 January 2024, there were 12,571 Mongolian legal documents. Moreover, (1) 150,375 criminal court orders—including 121,624 orders of a first instance, 24,142 orders of an appellate stage, and 4632 orders of a control stage; (2) 289,162 civil court orders—including 242,461 orders of a first instance, 33,660 orders of an appellate stage, and 13,041 orders of a control stage; and (3) 31,345 administrative court orders—including 18,174 orders of a first instance, 9150 orders of an appellate stage, and 4025 orders of a control stage can be found at the database of Mongolian court orders [75].

However, these legal documents have not been analyzed, which is mainly due to the lack of NLP tools that can handle Mongolian legal documents. Further computational analysis is therefore needed. It is a labor-intensive task to convert the above data to a machine-readable format and build gold standard corpora. On the other hand, in each year the Mongolian Bar Association [76] publishes a book that contains Mongolian bar exam questions and corresponding correct answers. The number of questions in 2016 was 4406; in 2019 was 4174; in 2021 was 4500; and in 2022 was 4598. Due to the unavailability of the properly labeled NLI dataset in the Mongolian legal domain, the authors' first contribution in the paper is the creation of an NLI dataset from Mongolian bar exam questions, which will be explained in Section 3.1.

3. NLI of Mongolian Bar Exam Questions

This section discusses the author’s approach to recognizing textual inference in Mongolian bar exam questions. The objective of recognizing textual inference in Mongolian bar exam questions is similar to NLI, and it predicts entailment between a given premise, i.e., a law or an article from a law, and the given hypothesis, which is a statement of a legal question. However, to tolerate the legal exactness, we do not use a “neutral” label. Thus, if the hypothesis entails the premise, the label is “True” (entailment), or if the hypothesis does not entail the premise, the label is “False” (contradiction).

3.1. An NLI Dataset of Mongolian Bar Exam Questions

In this research, an NLI dataset from Mongolian bar exam questions was prepared for recognizing textual inference in the Mongolian language. The books of the Mongolian bar exam questions have an average of 4500 questions with an enormous amount of content that requires expert knowledge. Despite the labor-intensive and time-consuming task, an NLI dataset was compiled manually by selecting 829 questions related to Mongolian civil law from the Mongolian bar exam questions. To reflect with the COLIEE, a legal entailment competition that was held over ten years; all 829 questions related to Mongolian civil law were chosen in this research among categories such as the constitution, human rights, criminal code, general administration, education, higher education, the central bank, health, health insurance, taxation, etc., of the Mongolian bar exams. The training data of the latest COLIEE-2023 contain 996 pairs of legal questions and Japanese civil law articles. A Mongolian bar exam question is a multiple choice test with four answers, including one correct answer and three incorrect answers. Each question was substituted with the corresponding law article by human experts in the Mongolian legal domain, and it was utilized as a “premise”. As the “hypothesis”, the 415 correct answers were utilized with the label “True”, and the 414 incorrect answers were utilized with the label “False”. After checking and understanding the contents of the four answers, human experts selected the corresponding articles in the Mongolian laws. The Civil Code articles are usually very detailed and require careful attention, which need to be understood one by one. Examples of an answer (hypothesis) to a Mongolian bar exam question and the corresponding article (premise) from Mongolian civil law are shown in Table 1. Please refer to Table A1 for more examples. This dataset was used in all experiments.

Table 1. Examples of NLI dataset of Mongolian bar exam questions.

Hypothesis	Premise	Label
Тухайн асуудлыг нарийвчлан зохицуулсан хуулийн, тийм хууль байхгүй бол сүүлд хүчин төгөлдөр болсон хуулийг хэрэглэнэ. (If there is a law that regulates the issue in detail, it will be used. Otherwise, if there is no such law, the latest law that came into force will be used. *)	3.3. Монгол Улсын Үндсэн хууль, энэ хуулиас бусад хууль хоорондоо зөрчилдвөл тухайн асуудлыг илүү нарийвчлан зохицуулсан хуулийн, тийм хууль байхгүй бол сүүлд хүчин төгөлдөр болсон хуулийн заалтыг хэрэглэнэ. (3.3. If laws, except the Constitution and Case Law, contradict each other, the issue will be resolved by the law that regulates it in more detail. If there is no such law, the provisions of the later effective law shall be applied. *)	True
Худалдаж байгаа эд хөрөнгийн талаар мэдээлэл өгөх нь үүрэг биш эрх учир хэн ч үүрэг хүлээхгүй. (The seller has an obligation to provide information about the goods for sale, because it is not an obligation but a right. *)	243.2. Худалдагч нь худалдаж байгаа эд хөрөнгийн зориулалт, хэрэглээний шинж чанар, хадгалах, хэрэглэх, тээвэрлэх нөхцөл, журам, баталгаат болон эдэлгээний хугацаа, үйлдвэрлэгчтэй холбогдсон үнэн зөв, бүрэн мэдээллийг худалдан авагчид өгөх үүрэгтэй. (243.2. The seller shall be obligated to provide the buyer with accurate and complete information about the designation, usage characteristics, storing, using, and transporting conditions and procedures, warranty and guarantee period, and the manufacturer of the goods sold. *)	False

* Unofficial English translation by the authors.

3.2. Language Modeling for Predicting NLI Labels in Mongolian Bar Exam Questions

Transformer-based language models were utilized for recognizing textual inference in Mongolian bar exam questions and predicting NLI labels. In other words, pretrained transformer models were fine-tuned in the Mongolian legal domain. The NLI was treated as a classification problem, which aims to recognize textual inference of hypothesis–premise pairs and label them as “entailment” or “contradiction”. We followed the standard practice for sentence pair tasks as in Devlin et al. [8]. Thus, “premise” and “hypothesis” were conjugated with a separate token [SEP], prepended to the “classification” token [CLS], and the sequence was input into the transformer models.

First, the experiments to predict textual inference in Mongolian bar exam questions using the existing pretrained models were conducted, which are explained in Section 3.3. Then, the existing pretrained models were fine-tuned for recognizing textual inference in Mongolian bar exam questions, and the achievements are discussed in Section 3.4. Detailed fine-tuning settings and hyperparameters are also introduced in Section 3.4.1.

3.3. The Performances of the Existing Pretrained Models in Recognizing Textual Inference in Mongolian Bar Exam Questions

Experiments were conducted to predict textual inference in the NLI dataset of Mongolian bar exam questions (explained in Section 3.1) using the existing pretrained models, including the (1) mongolian-roberta-base [50]; (2) mongolian-roberta-large [51]; (3) albert-mongolian [52]; (4) bert-base-mongolian-cased [55]; (5) bert-large-mongolian-uncased [56]; (6) bert-base-multilingual-cased [59]; (7) bert-large-mongolian-cased [77]; and (8) bert-base-mongolian-uncased [78]. Training was run only for the top layers to use the representations learned by existing pretrained models to extract features from new samples, i.e., the NLI dataset of Mongolian bar exam questions. A new classifier was added to label Mongolian bar exam questions’ pairs as “entailment” or “contradiction”. A total of 829 pairs of Mongolian bar exam questions were split through random shuffling with a training–validation–test split ratio of 80:10:10, respectively. During these experiments, the above pretrained models were used with the default settings, a training batch size of 16, and five epochs. The performance outcomes of the textual inference tasks utilizing the existing pretrained models are shown in Table 2. The best results are shown in bold text.

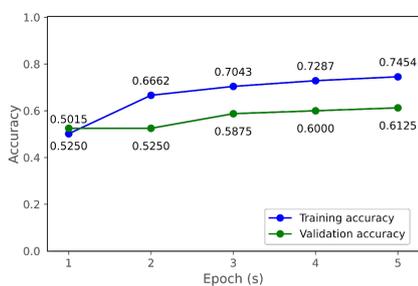
The bert-large-mongolian-uncased [56] obtained the highest average accuracy of 0.7381, whereas the mongolian-roberta-large [51] had the lowest average accuracy of 0.5357 in the unseen test data. Although the bert-base-mongolian-uncased [78] obtained an F1 score of 0.7294 in recognizing “entailment”, the bert-large-mongolian-uncased [56] obtained an F1 score of 0.7250 in recognizing “entailment” and an F1 score of 0.7500 in recognizing “contradiction”. In general, the existing pretrained bert-large-mongolian-uncased [56] model achieved the best performance in recognizing textual inference in the unseen test data.

As illustrated in Figure 1a–h, the training and validation accuracy improved incrementally after each epoch. In most cases, the validation accuracy was lower than the training accuracy.

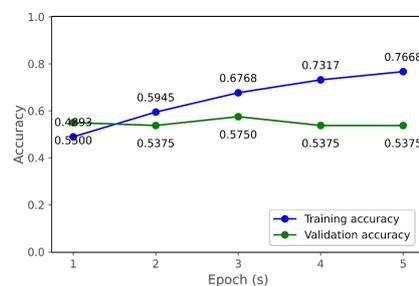
The performance outcomes of the fine-tuned models for recognizing textual inference in Mongolian bar exam questions are described in the next section. Detailed fine-tuning settings and hyperparameters are also introduced there.

Table 2. The performance outcomes of the existing models in recognizing textual inference in Mongolian bar exam questions.

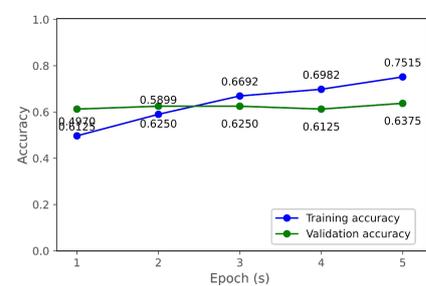
Label	Precision	Recall	F1 score	Label	Precision	Recall	F1 score
Contradiction	0.5738	0.8333	0.6796	Contradiction	0.5385	0.5000	0.5185
Entailment	0.6957	0.3810	0.4923	Entailment	0.5333	0.5714	0.5517
Accuracy			0.6071	Accuracy			0.5357
Macro average	0.6347	0.6071	0.5860	Macro average	0.5359	0.5357	0.5351
Weighted average	0.6347	0.6071	0.5860	Weighted average	0.5359	0.5357	0.5351
a mongolian-roberta-base [50]				b mongolian-roberta-large [51]			
Label	Precision	Recall	F1 score	Label	Precision	Recall	F1 score
Contradiction	0.5882	0.7143	0.6452	Contradiction	0.6600	0.7857	0.7174
Entailment	0.6364	0.5000	0.5600	Entailment	0.7353	0.5952	0.6579
Accuracy			0.6071	Accuracy			0.6905
Macro average	0.6123	0.6071	0.6026	Macro average	0.6976	0.6905	0.6876
Weighted average	0.6123	0.6071	0.6026	Weighted average	0.6976	0.6905	0.6876
c albert-mongolian [52]				d bert-base-mongolian-cased [55]			
Label	Precision	Recall	F1 score	Label	Precision	Recall	F1 score
Contradiction	0.7174	0.7857	0.7500	Contradiction	0.7500	0.5714	0.6486
Entailment	0.7632	0.6905	0.7250	Entailment	0.6538	0.8095	0.7234
Accuracy			0.7381	Accuracy			0.6905
Macro average	0.7403	0.7381	0.7375	Macro average	0.7019	0.6905	0.6860
Weighted average	0.7403	0.7381	0.7375	Weighted average	0.7019	0.6905	0.6860
e bert-large-mongolian-uncased [56]				f bert-base-multilingual-cased [59]			
Label	Precision	Recall	F1 score	Label	Precision	Recall	F1 score
Contradiction	0.7273	0.5714	0.6400	Contradiction	0.7317	0.7143	0.7229
Entailment	0.6471	0.7857	0.7097	Entailment	0.7209	0.7381	0.7294
Accuracy			0.6786	Accuracy			0.7262
Macro average	0.6872	0.6786	0.6748	Macro average	0.7263	0.7262	0.7262
Weighted average	0.6872	0.6786	0.6748	Weighted average	0.7263	0.7262	0.7262
g bert-large-mongolian-cased [77]				h bert-base-mongolian-uncased [78]			



(a) mongolian-roberta-base



(b) mongolian-roberta-large



(c) albert-mongolian

Figure 1. Cont.

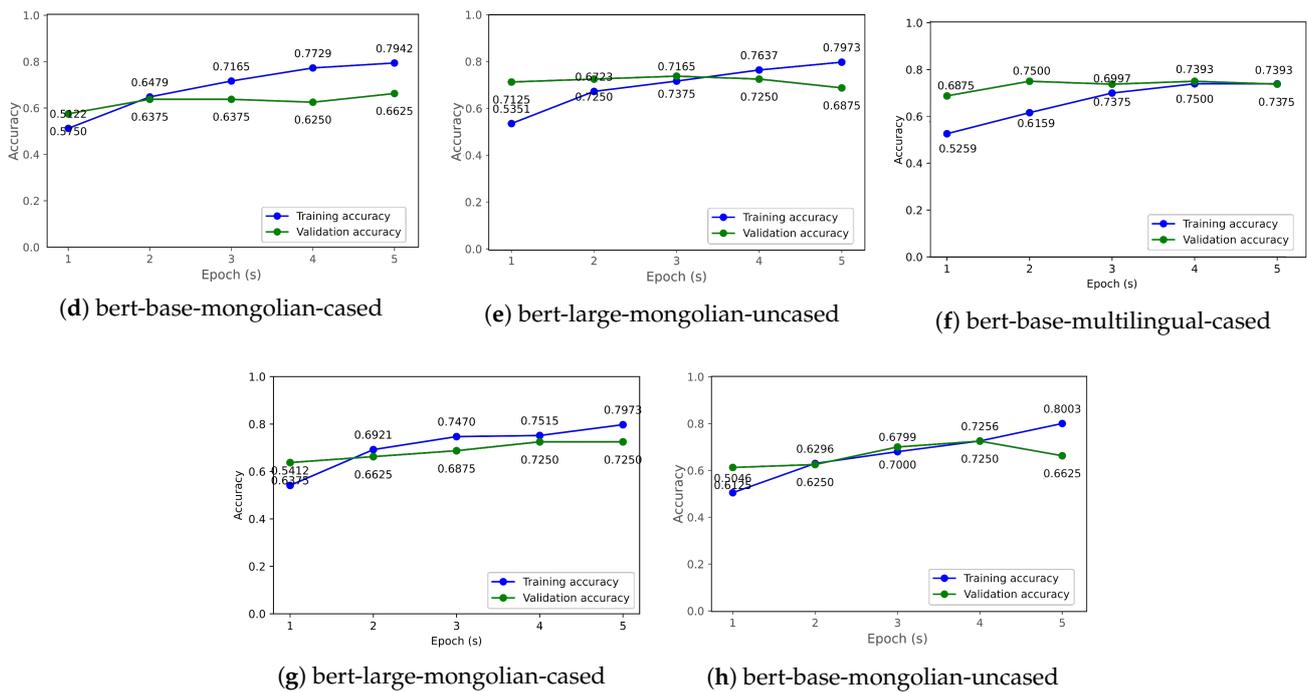


Figure 1. Training and validation accuracy for each epoch.

3.4. Fine-Tuning Pretrained Transformer Models in Recognizing Textual Inference in Mongolian Bar Exam Questions

The existing pretrained models introduced in Section 3.3 were fine-tuned for recognizing textual inference in Mongolian bar exam questions by unfreezing and retraining. The legal domain has specific vocabulary and characteristics in legal texts. Thus, as discussed in Section 3.3, the performance outcomes of the existing pretrained models in Mongolian legal texts were not decent. Fine-tuning allows us to adapt the feature representations in the existing pretrained models to the new samples, i.e., the NLI dataset of Mongolian bar exam questions for making the existing pretrained models more applicable to the Mongolian legal NLI task. The setup and dataset are described below.

3.4.1. Setup

The existing pretrained models were unfrozen and retrained with the following hyperparameters: a batch size of 16, a learning rate of 1×10^{-5} , a dropout rate of 0.3, a “softmax” activation function, and an Adam optimizer. Other experimental settings were the same as the experiments in Section 3.3. The classifier was also the same as the experiments in Section 3.3, which labeled Mongolian bar exam questions’ pairs as “entailment” or “contradiction”. All training was run for five epochs. Please refer to Table 3 for more details about each model.

Table 3. Settings of the existing pretrained models.

Model Name	Number of Parameters	Number of Layers	Vocabulary Size
mongolian-roberta-base [50]	125,072,642	12	50,265
mongolian-roberta-large [51]	355,917,826	24	50,265
albert-mongolian [52]	12,110,594	12	30,000
bert-base-mongolian-cased [55]	111,044,354	12	32,000
bert-large-mongolian-uncased [56]	337,213,442	24	32,000
bert-base-multilingual-cased [59]	177,853,440	12	32,000
bert-large-mongolian-cased [77]	337,213,442	24	32,000
bert-base-mongolian-uncased [78]	111,044,354	12	32,000

3.4.2. Datasets

The same training, validation, and test data of 829 pairs of Mongolian bar exam questions that were split with a training–validation–test split ratio of 80:10:10, respectively, were used in all experiments. The experimental data distribution of the Mongolian bar exam questions is shown in Table 4.

Table 4. Experimental data distribution.

Dataset	Entailment	Contradiction	Total
Training data	332	331	663
Validation data	41	41	82
Test data	42	42	84
Total	415	414	829

The maximum tokens of the pairs of Mongolian bar exam questions, which were determined by the tokenizer of each pretrained model, are shown in Table 5.

Table 5. The length of the maximum token sequence.

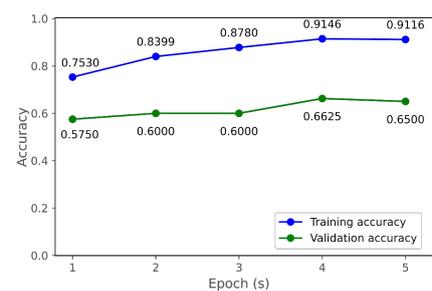
Model Name	Maximum Tokens
mongolian-roberta-base [50]	268
mongolian-roberta-large [51]	268
albert-mongolian [52]	270
bert-base-mongolian-cased [55]	251
bert-large-mongolian-uncased [56]	249
bert-base-multilingual-cased [59]	512
bert-large-mongolian-cased [77]	251
bert-base-mongolian-uncased [78]	249

4. Experimental Results of Recognizing Textual Inference in Mongolian Bar Exam Questions

The performance outcomes of the textual inference tasks utilizing the fine-tuned models are shown in Figures 2–9.

Label	Precision	Recall	F1 score
Contradiction	0.7297	0.6429	0.6835
Entailment	0.6809	0.7619	0.7191
Accuracy			0.7024
Macro average	0.7053	0.7024	0.7013
Weighted average	0.7053	0.7024	0.7013

(a)



(b)

Figure 2. Performance outcomes of the fine-tuned mongolian-roberta-base. (a) Evaluation metrics: Precision, Recall, F1 score and Accuracy. (b) Training and validation accuracy for each epoch.

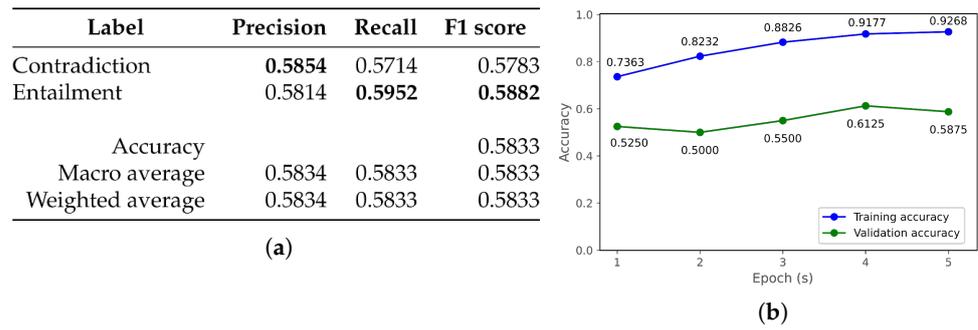


Figure 3. Performance outcomes of the fine-tuned mongolian-roberta-large. (a) Evaluation metrics: Precision, Recall, F1 score and Accuracy. (b) Training and validation accuracy for each epoch.

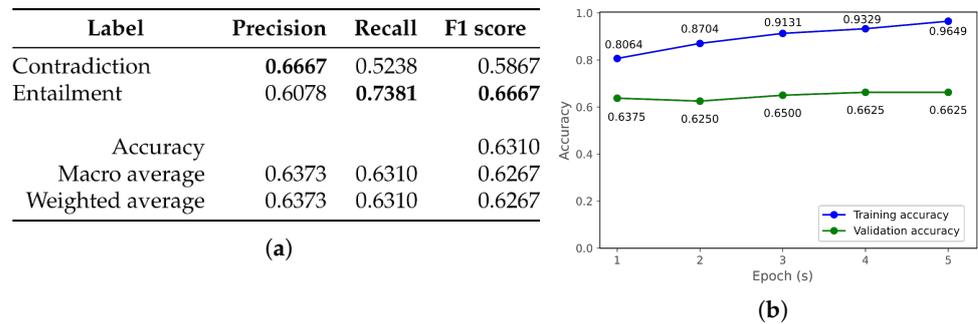


Figure 4. Performance outcomes of the fine-tuned albert-mongolian. (a) Evaluation metrics: Precision, Recall, F1 score and Accuracy. (b) Training and validation accuracy for each epoch.

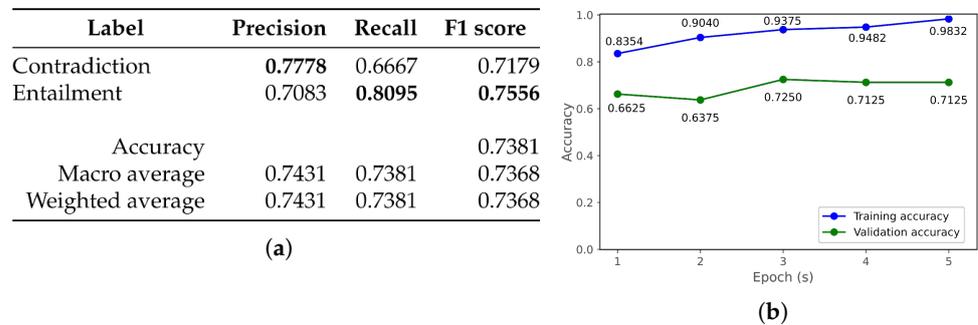


Figure 5. Performance outcomes of the fine-tuned bert-base-mongolian-cased. (a) Evaluation metrics: Precision, Recall, F1 score and Accuracy. (b) Training and validation accuracy for each epoch.

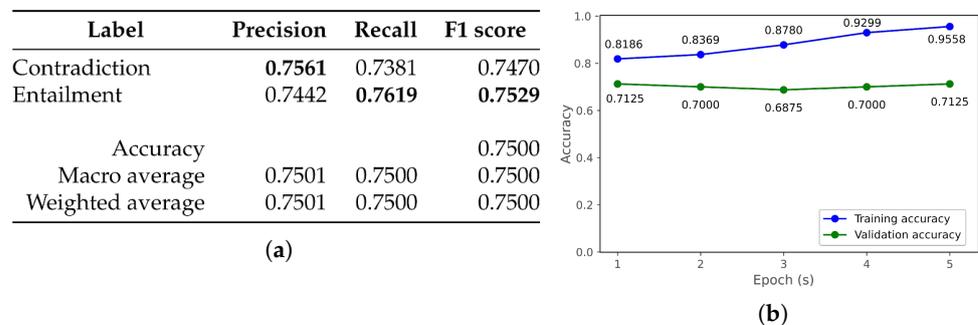


Figure 6. Performances of the fine-tuned bert-large-mongolian-uncased. (a) Evaluation metrics: Precision, Recall, F1 score and Accuracy. (b) Training and validation accuracy for each epoch.

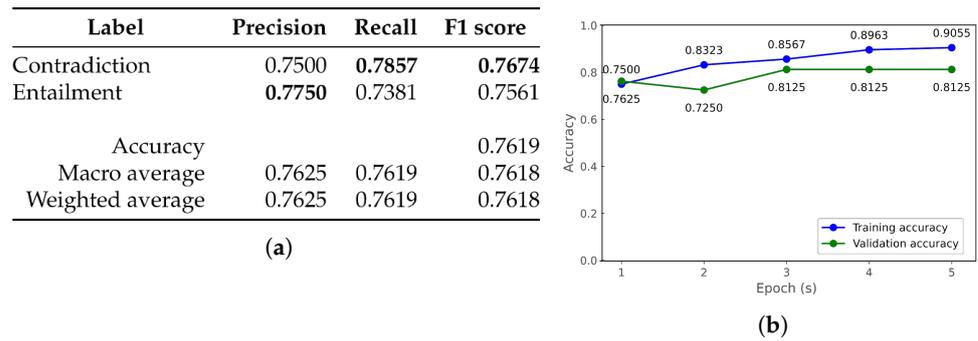


Figure 7. Performance outcomes of the fine-tuned bert-base-multilingual-cased. (a) Evaluation metrics: Precision, Recall, F1 score and Accuracy. (b) Training and validation accuracy for each epoch.

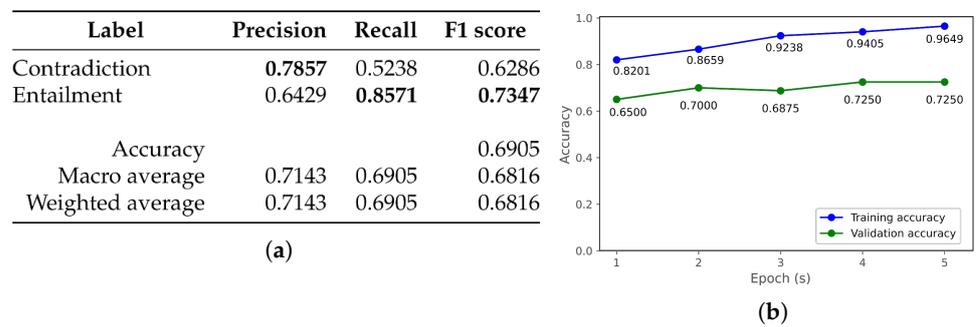


Figure 8. Performance outcomes of the fine-tuned bert-large-mongolian-cased. (a) Evaluation metrics: Precision, Recall, F1 score and Accuracy. (b) Training and validation accuracy for each epoch.

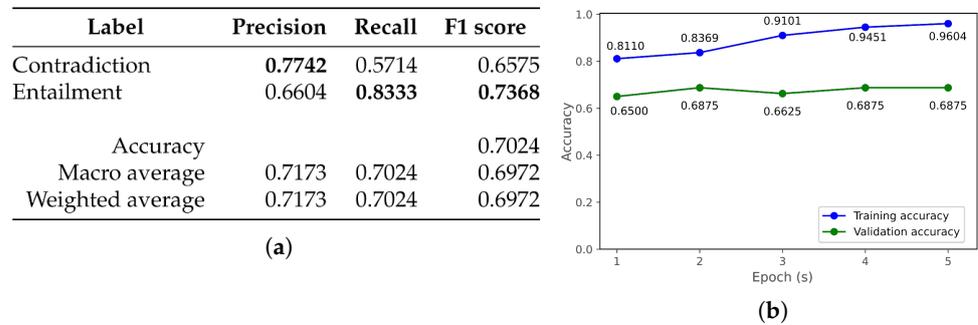


Figure 9. Performance outcomes of the fine-tuned bert-base-mongolian-uncased. (a) Evaluation metrics: Precision, Recall, F1 score and Accuracy. (b) Training and validation accuracy for each epoch.

Among the eight different models, the fine-tuned bert-base-multilingual-cased achieved the highest average accuracy of 0.7619, the best F1 score of 0.7561 in recognizing “entailment”, and the best F1 score of 0.7674 in recognizing “contradiction” in the unseen test data. The highest recall of 0.8571 (See Figure 8a) was obtained in recognizing “entailment” using the fine-tuned bert-large-mongolian-cased. Moreover, the highest precision of 0.7857 (See Figure 8a) was obtained in recognizing “contradiction” using the fine-tuned bert-large-mongolian-cased. On the contrary, as shown in Figure 6a, the fine-tuned bert-large-mongolian-uncased demonstrated balanced performance in recognizing textual inference in Mongolian bar exam questions, thus achieving a precision of 0.7561, a recall of 0.7381, and an F1 score of 0.7470 for recognizing “contradiction”. It also achieved a precision of 0.7442, a recall of 0.7619, and an F1 score of 0.7529 for recognizing “entailment.” In contrast, the fine-tuned mongolian-roberta-large performed less successfully, thereby having the lowest average accuracy of 0.5833 in the unseen test data. As shown in Figure 3a, the fine-tuned mongolian-roberta-large lagged behind, thereby obtaining the lowest precision, recall, and

F1 score in recognizing both the “contradiction” and “entailment” categories. It obtained the lowest precision of 0.5854, recall of 0.5714, and F1 score of 0.5783 for “contradiction,” as well as a precision of 0.5814, recall of 0.5952, and an F1 score of 0.5882 for “entailment”.

An overall comparison in recognizing textual inference in Mongolian bar exam questions using different models is shown in Table 6 with an accuracy, macro average F1 score, and weighted average F1 score. Table 6 also compares the performance of fine-tuned models against the existing pretrained models. The fine-tuned bert-base-multilingual-cased [59] model showed an average accuracy of 0.7619, a macro average F1 score of 0.7618, and a weighted average F1 score of 0.7618. In general, the fine-tuned bert-base-multilingual-cased model achieved the best performance in recognizing textual inference in Mongolian bar exam questions.

Table 6. Performance comparison in recognizing textual inference in Mongolian bar exam questions using different models.

Model Name	Accuracy		F1 Score			
			Macro Average		Weighted Average	
	Before *	After **	Before *	After **	Before *	After **
mongolian-roberta-base [50]	0.6071	0.7024	0.5860	0.7013	0.5860	0.7013
mongolian-roberta-large [51]	0.5357	0.5833	0.5351	0.5833	0.5351	0.5833
albert-mongolian [52]	0.6071	0.6310	0.6026	0.6267	0.6026	0.6267
bert-base-mongolian-cased [55]	0.6905	0.7381	0.6876	0.7368	0.6876	0.7368
bert-large-mongolian-uncased [56]	0.7381	0.7500	0.7375	0.7500	0.7375	0.7500
bert-base-multilingual-cased [59]	0.6905	0.7619	0.6860	0.7618	0.6860	0.7618
bert-large-mongolian-cased [77]	0.6786	0.6905	0.6748	0.6816	0.6748	0.6816
bert-base-mongolian-uncased [78]	0.7262	0.7024	0.7262	0.6972	0.7262	0.6972

* The performance of the existing pretrained models. ** The performance of the fine-tuned models.

As illustrated in Figures 2b, 3b, 4b, 5b, 6b, 7b, 8b, and 9b, the training and validation accuracy improved incrementally after each epoch. The training accuracy ranged from 0.7363 to 0.9832, while the validation accuracy ranged from 0.5000 to 0.8125.

The confusion matrices of the fine-tuned models on unseen test data are shown in Figure 10. The numbers inside a bracket represent the percentage within the total test data. As illustrated in Figure 10f, the fine-tuned bert-base-multilingual-cased incorrectly labeled 21.42% of the “contradiction” pairs in the test data as “entailment” in recognizing textual inference in Mongolian bar exam questions. Also, as shown in Figure 10g, in the fine-tuned bert-large-mongolian-cased, 14.28% of “entailment” pairs were incorrectly labeled as “contradiction”. Overall, the fine-tuned bert-base-multilingual-cased showed a better performance in recognizing textual inference in Mongolian bar exam questions.

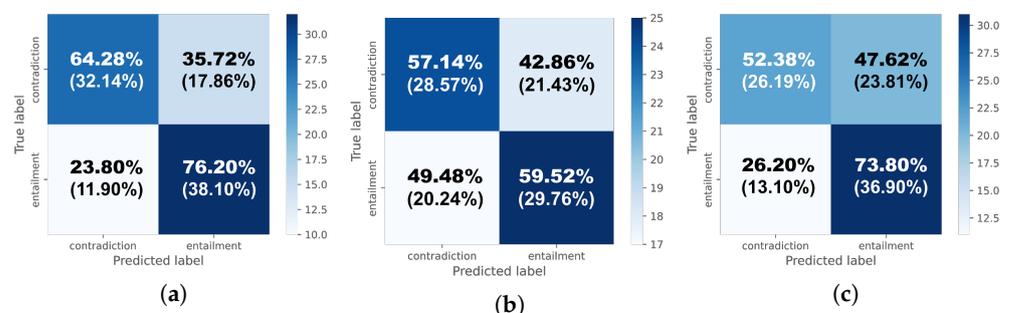


Figure 10. Cont.

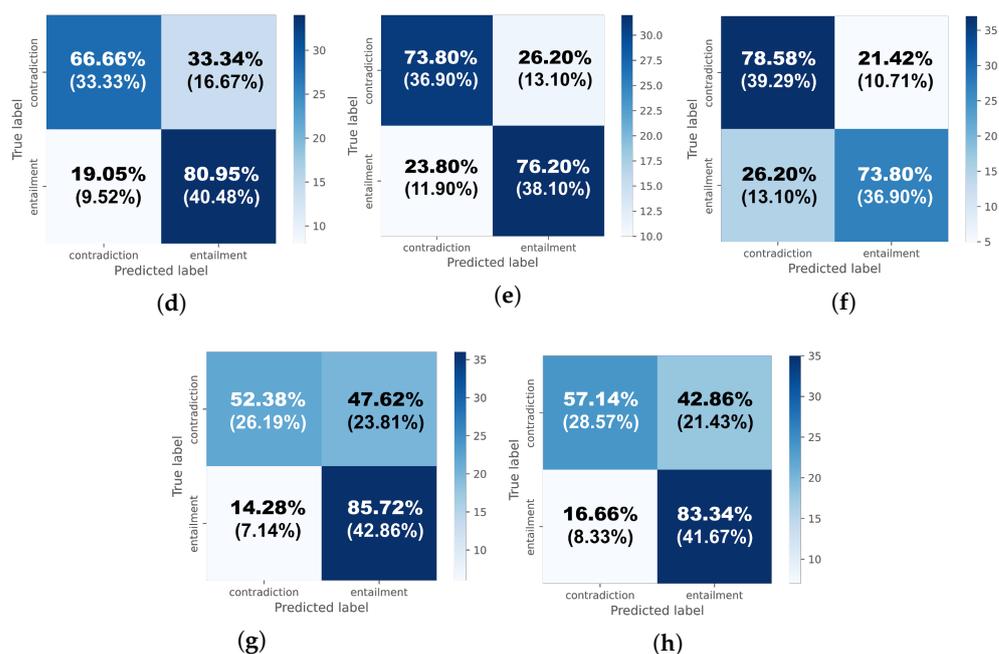


Figure 10. Confusion matrices of the fine-tuned models over the test data in textual inference tasks. (a) the fine-tuned mongolian-roberta-base, (b) the fine-tuned mongolian-roberta-large, (c) the fine-tuned albert-mongolian, (d) the fine-tuned bert-base-mongolian-cased, (e) the fine-tuned bert-large-mongolian-uncased, (f) the fine-tuned bert-base-multilingual-cased, (g) the fine-tuned bert-large-mongolian-cased, (h) the fine-tuned bert-base-mongolian-uncased.

5. Conclusions

In this paper, the existing deep learning models were examined for recognizing textual inference in Mongolian bar exam questions. Several fine-tuned transformer-based models were investigated, which are important for the DSS that we aim to develop. The demonstrated fine-tuned models were evaluated in recognizing textual inference in Mongolian bar exam questions. Overall, as shown in Table 6, the fine-tuned bert-base-multilingual-cased [59] model showed the best results in recognizing textual inference in Mongolian bar exam questions. It was capable of recognizing “contradiction” with a precision of 0.7500, a recall of 0.7857, and an F1 score of 0.7674, as well as recognizing “entailment” with a precision of 0.7750, a recall of 0.7381, and an F1 score of 0.7561. The demo system has been developed, and it can be accessed online at https://www.dl.is.ritsumei.ac.jp/legal_analysis/NLI.html (accessed on 18 January 2024).

In future work, some distinct features need to be investigated to improve the accuracy of distinguishing “contradiction” and “entailment” more accurately. The Mongolian bar exam questions may contain many common or similar sentences. The positive and negative sampling methods or data augmentation need to be considered for further improvements. Our further research will apply LLMs to identify conflicting Mongolian legal texts.

Author Contributions: Conceptualization, G.K., B.B. and A.M.; methodology, G.K.; software, G.K. and B.B.; validation, G.K. and B.B.; formal analysis, G.K.; investigation, G.K.; resources, G.K. and B.B.; data curation, G.K. and B.B.; writing—original draft preparation, G.K.; writing—review and editing, B.B. and A.M.; visualization, G.K. and B.B.; supervision, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by JSPS KAKENHI Grant Number 21K12600.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on request from the authors. The raw data supporting the conclusions of this article will be made available by the authors on request. The demo system is available online at https://www.dl.is.ritsumei.ac.jp/legal_analysis/NLI.html (accessed on 18 January 2024).

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

Appendix A

As supplementary data, Table A1 lists some more examples of the NLI dataset of Mongolian bar exam questions.

Table A1. Examples of the NLI dataset of Mongolian bar exam questions.

Hypothesis	Premise	Label
Хуульд зааснаар худалдааны төлөөлөгчид олгох нөхөн олговор нь түүний үйл ажиллагааны сүүлийн 5 жилд авч байсан, эсхүл 1 жилд олох хөлс, шагналын дундаж хэмжээнээс илүүгүй байна. (According to the law, the compensation to be given to the sales representative should not be more than the average revenue earned in the last 5 years of his/her activity or expected to be earned in 1 year. *)	418.3. Нөхөн олговор нь худалдааны төлөөлөгчийн үйл ажиллагааны сүүлийн таван жилд авч байсан, эсхүл нэг жилд олох хөлс, шагналын дундаж хэмжээнээс илүүгүй байна. Нэг жилээс доош хугацаатай байгуулсан гэрээнд худалдааны төлөөлөгчийн үйл ажиллагааны хугацаанд олж болох хөлс, шагналын дундаж дүнгээс тооцно. (418.3. The compensation shall not exceed the average salary or bonus earned in the last five years or one year’s earnings of the sales representative’s activity. For contracts lasting a period of less than one year, the compensation shall be calculated from the average income that can be earned during the period of activity of the sales representative. *)	True
Төрөөгүй байгаа хүүхдэд эд хөрөнгө хуваарилагдахгүй. (Property will not be distributed to unborn children. *)	532.2. Өвлүүлэгчийг амьд байхад олдсон бөгөөд төрөөгүй байгаа өвлөгчид оногдох хэсгийг тусгаарлан гаргана. (532.2. The portion of the inheritance to the unborn heir, who has been a fetus while the testator was alive, shall be treated separately. *)	False
Хуульд зааснаар өвлөгчдийн хооронд үүссэн маргааныг шүүх хүлээн авч шийдвэрлэх боломжгүй. (According to the law, disputes between heirs cannot be accepted and resolved by the court. *)	532.1. Өвлөгдсөн эд хөрөнгийг өв залгамжлалд оролцсон бүх өвлөгчид хэлэлцэн зөвшөөрөлцөж, хууль ёсны буюу гэрээслэлээр өвлөгч бүрт оногдвол зохих хэмжээгээр хуваарилах бөгөөд энэ талаар маргаан гарвал шүүх шийдвэрлэнэ. (532.1. If it is assigned to each heir legally or by will, the inherited property shall be distributed according to the appropriate amount after all the heirs participating in the inheritance have approved it, and any disputes shall be resolved by the court. *)	False

* Unofficial English translation by the authors.

References

- World Development Indicators. Available online: <https://databank.worldbank.org/source/world-development-indicators> (accessed on 9 January 2024).
- Mid-Term Evaluation’s Discussion of the Medium-Term Action Plan “New Development”. Available online: <https://parliament.mn/n/gico> (accessed on 9 January 2024).
- Shagdarsuren, T. *Study of Mongolian Scripts (Graphic Study of Grammatology*, 2nd ed.; Urlakh Erdem Khevreliin Gazar: Ulaanbaatar, Mongolia, 2001; pp. 3–11.
- Svantesson, J.; Tsendina, A.; Karlsson, A.; Franzén, V. *The phonology of Mongolian*, 1st ed.; Oxford University Press: New York, NY, USA, 2005; pp. xv–xix.
- Hirschberg, J.; Manning, C.D. Advances in natural language processing. *Science* **2015**, *349*, 261–266. [[CrossRef](#)] [[PubMed](#)]
- Pouyanfar, S.; Sadiq, S.; Yan, Y.; Tian, H.; Tao, Y.; Reyes, M.P.; Shyu, M.; Chen, S.; Iyengar, S.S. A survey on deep learning: Algorithms, techniques, and applications. *ACM Comput. Surv.* **2018**, *51*, 1–36. [[CrossRef](#)]

7. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language processing [Review article]. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75. [[CrossRef](#)]
8. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
9. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2020**, *21*, 1–67.
10. Bowman, S.R.; Angeli, G.; Potts, C.; Manning, C.D. A large annotated corpus for learning natural language inference. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 632–642.
11. Williams, A.; Nangia, N.; Bowman, S. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), New Orleans, LA, USA, 1–6 June 2018; pp. 1112–1122.
12. Introducing ChatGPT. Available online: <https://openai.com/blog/chatgpt> (accessed on 18 January 2024).
13. Competition on Legal Information Extraction/Entailment (COLIEE). Available online: <https://sites.ualberta.ca/~rabelo/COLIEE2023/> (accessed on 18 January 2024).
14. Yoshioka, M.; Suzuki, Y.; Aoki, Y. HUKB at the COLIEE 2022 Statute Law Task. In *New Frontiers in Artificial Intelligence. JSAI-isAI 2022*; Takama, Y., Yada, K., Satoh, K., Arai, S., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2022; Volume 13859, pp. 109–124.
15. Yoshioka, M.; Aoki, Y. HUKB at COLIEE 2023 Statute Law Task. In Proceedings of the Tenth Competition on Legal Information Extraction/Entailment (COLIEE'2023), Braga, Portugal, 19 June 2023; pp. 72–76.
16. Fujita, M.; Onaga, T.; Ueyama, A.; Kano, Y. Legal Textual Entailment Using Ensemble of Rule-Based and BERT-Based Method with Data Augmentation by Related Article Generation. In *New Frontiers in Artificial Intelligence. JSAI-isAI 2022*; Takama, Y., Yada, K., Satoh, K., Arai, S., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2022; Volume 13859, pp. 138–153.
17. Bui, Q.; Do, T.; Le, K.; Nguyen, D.H.; Nguyen, H.; Pham, T.; Nguyen, L. JNLP COLIEE-2023: Data Argumentation and Large Language Model for Legal Case Retrieval and Entailment. In Proceedings of the Tenth Competition on Legal Information Extraction/Entailment (COLIEE'2023), Braga, Portugal, 19 June 2023; pp. 17–26.
18. Chung, H.W.; Hou, L.; Longpre, S.; Zoph, B.; Tay, Y.; Fedus, W.; Li, Y.; Wang, X.; Dehghani, M.; Brahma, S.; et al. Scaling instruction-finetuned language models. *arXiv* **2023**, arXiv:2210.11416.
19. Tay, Y.; Dehghani, M.; Tran, V.Q.; Garcia, X.; Wei, J.; Wang, X.; Chung, H.; Shakeri, S.; Bahri, D.; Schuster, T.; et al. UL2: Unifying Language Learning Paradigms. *arXiv* **2023**, arXiv:2205.05131.
20. Flan-Alpaca: Instruction Tuning from Humans and Machines. Available online: <https://huggingface.co/declare-lab/flan-alpaca-xxl> (accessed on 18 January 2024).
21. Onaga, T.; Fujita, M.; Kano, Y. Japanese Legal Bar Problem Solver Focusing on Person Names. In Proceedings of the Tenth Competition on Legal Information Extraction/Entailment (COLIEE'2023), Braga, Portugal, 19 June 2023; pp. 63–71.
22. Luke-Japanese. Available online: <https://huggingface.co/studio-ousia/luke-japanese-base-lite> (accessed on 18 January 2024).
23. Nguyen, C.; Nguyen, P.; Tran, T.; Nguyen, D.; Trieu, A.; Pham, T.; Dang, A.; Nguyen, M. CAPTAIN at COLIEE 2023: Efficient Methods for Legal Information Retrieval and Entailment Tasks. In Proceedings of the Tenth Competition on Legal Information Extraction/Entailment (COLIEE'2023), Braga, Portugal, 19 June 2023; pp. 7–16.
24. Nogueira, R.; Jiang, Z.; Pradeep, R.; Lin, J. Document Ranking with a Pretrained Sequence-to-Sequence Model. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020, Online, 16–20 November 2020; pp. 708–718.
25. Ariunaa, O.; Munkhjargal, Z. Sentiment Analysis for Mongolian Tweets with RNN. In Proceedings of the 16th International Conference on IJHMSP in Conjunction with the 13th International Conference on FITAT, Ho Chi Minh City, Vietnam, 5–7 November 2020; pp. 74–81.
26. Battumur, K.; Dulamragchaa, U.; Enkhbat, S.; Altankhuyag, L.; Tumurbaatar, P. Cyrillic Word Error Program Based on Machine Learning. *J. Inst. Math. Digit. Technol.* **2022**, *4*, 54–60. [[CrossRef](#)]
27. Dashdorj, Z.; Munkhbayar, T.; Grigorev, S. Deep learning model for Mongolian Citizens' Feedback Analysis using Word Vector Embeddings. *arXiv* **2023**, arXiv:2302.12069.
28. Choi, S.; Tsend, G. A Study on the Appropriate Size of the Mongolian General Corpus. *Int. J. Nat. Lang. Comput.* **2023**, *12*, 17–30.
29. Lkhagvasuren, G.; Rentsendorj, J.; Namsrai, O.E. Mongolian Part-of-Speech Tagging with Neural Networks. In Proceedings of the 16th International Conference on IJHMSP in Conjunction with the 13th International Conference on FITAT, Ho Chi Minh City, Vietnam, 5–7 November 2020; pp. 109–115.
30. Jaimai, P.; Chimeddorj, O. Part of speech tagging for Mongolian corpus. In Proceedings of the 7th Workshop on Asian Language Resources (ACL-IJCNLP2009), Suntec, Singapore, 6–7 August 2009; pp. 103–106.
31. Ivanov, B.; Musa, M.; Dulamragchaa, U. Mongolian Spelling Error Correction Using Word N-Gram Method. In Proceedings of the 16th International Conference on IJHMSP in Conjunction with the 13th International Conference on FITAT, Ho Chi Minh City, Vietnam, 5–7 November 2020; pp. 94–101.

32. Batsuren, K.; Ganbold, A.; Chagnaa, A.; Giunchiglia, F. Building the Mongolian Wordnet. In Proceedings of the 10th Global Wordnet Conference, Wroclaw, Poland, 23–27 July 2019; pp. 238–244.
33. Bataa, B.; Altangerel, K. Word sense disambiguation in Mongolian language. In Proceedings of the 7th International Forum on Strategic Technology (IFOST), Tomsk, Russia, 18–21 September 2012; pp. 1–4.
34. Jaimai, P.; Zundui, T.; Chagnaa, A.; Ock, C.Y. PC-KIMMO-based description of Mongolian morphology. *Int. J. Inf. Process. Syst.* **2005**, *1*, 41–48. [[CrossRef](#)]
35. Dulamragchaa, U.; Chadraabal, S.; Ivanov, B.; Baatarhuu, M. Mongolian language morphology and its database structure. In Proceedings of the International Conference Green Informatics (ICGI2017), Fuzhou, China, 15–17 August 2017; pp. 282–285.
36. Chagnaa, A.; Adiyatseren, B. Two Level Rules for Mongolian Language. In Proceedings of the 7th International Conference Multimedia, Information Technology and its Applications (MITA2011), Ulaanbaatar, Mongolia, 6–9 July 2011; pp. 130–133.
37. Munkhjargal, Z.; Chagnaa, A.; Jaimai, P. Morphological Transducer for Mongolian. In Proceedings of the International Conference Computational Collective Intelligence (ICCCI 2016), Halkidiki, Greece, 28–30 September 2016; pp. 546–554.
38. Enkhbayar, S.; Utsuro, T.; Sato, S. Mongolian Phrase Generation and Morphological Analysis based on Phonological and Morphological Constraints. *J. Nat. Lang. Process.* **2005**, *12*, 185–205. (In Japanese) [[CrossRef](#)] [[PubMed](#)]
39. Khaltar, B.O.; Fujii, A. A lemmatization method for Mongolian and its application to indexing for information retrieval. *Inf. Process. Manag.* **2009**, *45*, 438–451. [[CrossRef](#)]
40. Nyandag, B.E.; Li, R.; Indruska, G. Performance Analysis of Optimized Content Extraction for Cyrillic Mongolian Learning Text Materials in the Database. *J. Comp. Commun.* **2016**, *4*, 79–89. [[CrossRef](#)]
41. Munkhjargal, Z.; Bella, G.; Chagnaa, A.; Giunchiglia, F. Named Entity Recognition for Mongolian Language. In Proceedings of the International Conference Text, Speech, and Dialogue (TSD2015), Pilsen, Czech Republic, 14–17 September 2015; pp. 243–251.
42. Khaltar, B.O.; Fujii, A.; Ishikawa, T. Extracting Loanwords from Mongolian Corpora and Producing a Japanese-Mongolian Bilingual Dictionary. In Proceedings of the 21st International Conference Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney, Australia, 17–18 July 2006; pp. 657–664.
43. Khaltar, B.O.; Fujii, A. Extracting Loanwords from Modern Mongolian Corpora and Producing a Japanese-Mongolian Bilingual Dictionary. *J. Inf. Process. Soc. Japan* **2008**, *49*, 3777–3788.
44. Lkhagvasuren, G.; Rentsendorj, J. Open Information Extraction for Mongolian Language. In Proceedings of the 16th International Conference on IJHMSP in conjunction with the 13th International Conference on FITAT, Ho Chi Minh City, Vietnam, 5–7 November 2020; pp. 299–304.
45. Damiran, Z.; Altangerel, K. Author Identification: An Experiment based on Mongolian Literature using Decision Tree. In Proceedings of the 7th International Conference Ubi-Media Computing and Workshops (UMEDIA), Ulaanbaatar, Mongolia, 12–14 July 2014; pp. 186–189.
46. Damiran, Z.; Altangerel, K. Text classification experiments on Mongolian language. In Proceedings of the 8th International Forum on Strategic Technology (IFOST2013), Ulaanbaatar, Mongolia, 28 June– 1 July 2013; pp. 145–148.
47. Ehara, T.; Hayata, S.; Kimura, N. Mongolian to Japanese Machine Translation System. *J. Yamanashi Eiwa Coll.* **2011**, *9*, 27–40.
48. Enkhbayar, S.; Utsuro, T.; Sato, S. Japanese-Mongolian Machine Translation of Functional Expressions. In Proceedings of the 10th Annual Conference of Japanese Association for Natural Language Processing (NLP2004), Tokyo, Japan, 16–18 March 2004; B1-1. (In Japanese)
49. BERT Pretrained Models on Mongolian Datasets. Available online: <https://github.com/tugstugi/mongolian-bert/> (accessed on 18 January 2024).
50. Mongolian RoBERTa Base. Available online: <https://huggingface.co/bayartsogt/mongolian-roberta-base> (accessed on 18 January 2024).
51. Mongolian RoBERTa Large. Available online: <https://huggingface.co/bayartsogt/mongolian-roberta-large> (accessed on 18 January 2024).
52. ALBERT Pretrained Model on Mongolian Datasets. Available online: <https://github.com/bayartsogt-ya/albert-mongolian/> (accessed on 18 January 2024).
53. Mongolian GPT2. Available online: <https://huggingface.co/bayartsogt/mongolian-gpt2> (accessed on 18 January 2024).
54. Mongolian Text Classification. Available online: <https://github.com/sharavsambuu/mongolian-text-classification> (accessed on 18 January 2024).
55. BERT-BASE-MONGOLIAN-CASED. Available online: <https://huggingface.co/tugstugi/bert-base-mongolian-cased> (accessed on 18 January 2024).
56. BERT-LARGE-MONGOLIAN-UNCASED. Available online: <https://huggingface.co/tugstugi/bert-large-mongolian-uncased> (accessed on 18 January 2024).
57. mongolian-bert-ner. Available online: <https://github.com/enod/mongolian-bert-ner> (accessed on 18 January 2024).
58. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised Cross-lingual Representation Learning at Scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 8440–8451.
59. BERT. Available online: <https://github.com/google-research/bert/> (accessed on 18 January 2024).
60. The Mongolian Wordnet (MonWN). Available online: <https://github.com/kbatsuren/monwn> (accessed on 18 January 2024).
61. Khalkha Mongolian (khk) Paradigms. Available online: <https://github.com/unimorph/khk> (accessed on 18 January 2024).

62. MorphyNet. Available online: <https://github.com/kbatsuren/MorphyNet> (accessed on 18 January 2024).
63. Eduge.mn Dataset. Available online: <https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/eduge.csv.gz> (accessed on 18 January 2024).
64. News.rar. Available online: <https://disk.yandex.ru/d/z5e3MVnKvFvF6w> (accessed on 18 January 2024).
65. CC-100: Monolingual Datasets from Web Crawl Data. Available online: <https://data.statmt.org/cc-100/> (accessed on 18 January 2024).
66. Mongolian Government Agency—11-11.mn Dataset. Available online: <https://www.kaggle.com/datasets/enqush/mongolian-government-agency-1111mn-dataset> (accessed on 18 January 2024).
67. Mongolian NER Dataset. Available online: https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/NER_v1.0.json.gz (accessed on 18 January 2024).
68. Mongolian_personal_names.csv.gz. Available online: https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/mongolian_personal_names.csv.gz (accessed on 18 January 2024).
69. Mongolian_clan_names.csv.gz. Available online: https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/mongolian_clan_names.csv.gz (accessed on 18 January 2024).
70. Mongolian_company_names.csv.gz. Available online: https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/mongolian_company_names.csv.gz (accessed on 18 January 2024).
71. Districts.csv. Available online: <https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/districts.csv> (accessed on 18 January 2024).
72. Countries.csv. Available online: <https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/countries.csv> (accessed on 18 January 2024).
73. Mongolian_abbreviations.csv. Available online: https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/mongolian_abbreviations.csv (accessed on 18 January 2024).
74. Unified Legal Information System. Available online: <https://legalinfo.mn/> (accessed on 18 January 2024).
75. The Database of Mongolian Court Orders. Available online: <https://shuukh.mn/> (accessed on 18 January 2024).
76. Mongolian Bar Association. Available online: <https://www.mglbar.mn/index> (accessed on 18 January 2024).
77. BERT-LARGE-MONGOLIAN-CASED. Available online: <https://huggingface.co/tugstugi/bert-large-mongolian-cased> (accessed on 18 January 2024).
78. BERT-BASE-MONGOLIAN-UNCASED. Available online: <https://huggingface.co/tugstugi/bert-base-mongolian-uncased> (accessed on 18 January 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.