A Linguistics-based Deep Learning Approach to ETL for Automated Translation of English Language Data

Gomathi R D[1], Shanthi R[2], Mythili M[3] & Prabha K[4]

[1] Department of English, Kongu Engineering College, Erode, Tamil Nadu, India. E-mail: gomathimaheswaran6@gmail.com

[2] Department of English, Paavai Engineering College, Tamilnadu, Namakkal, India. E-mail: drshanthi14@gmail.com

[3] Department of English, Nandha Engineering College, Erode, India. E-mail: murugesanmythili@gmail.com

[4] Department of English, Kongu Arts and Science College (Autonomous), Erode - 638107, India. E-mail: prabhaenglish85@gmail.com

Correspondence: Gomathi R D, Department of English, Kongu Engineering College, Erode, Tamil Nadu-638060, India. E-mail: gomathimaheswaran6@gmail.com

Received: January 1, 2025 Accepted: March 12, 2025 Online Published: April 17, 2025

doi:10.5430/wjel.v15n5p362 URL: https://doi.org/10.5430/wjel.v15n5p362

Abstract

Automated translation of information regarding English language has become imperative in this worldwide phenomenon. The erstwhile methods of performing Extract, Transform, and Load (ETL) of data involve a lot of manual effort besides taking a long time and being error-prone. A linguistics-based deep learning approach is proposed to improve the efficiency and accuracy of ETL for automated translation of English language data. The particular approach is focused on the adoption of deep learning techniques for automatically processing and translating data in English language without extensive manual intervention. In conjunction with this, it provides the use of linguistics knowledge, such as syntax, semantics, and grammar, for building an accurate deep learning model for the extract-transform-load process of being applied in the translation process. The method has also proved to be promising in experimentation with results competing favorably with common ETL methods speed and accuracy and showing good scalability. Linguistic insights in combination with deep learning have created for the specified approach the possibility of bringing efficiency, accuracy, and automation to the translation of the English language.

Keywords: automation, translation, extract, transform, load, English Language Learners (ELL), deep learning

1. Introduction

Machine translation, or automated translation, refers to translation via computer algorithms that automatically translate text from one language into another without any human involvement. Moreover, the growth of machine translation in recent years has been very promising due to advances occurring mostly in Natural Language Processing (NLP) and Artificial Intelligence (AI) technology (Seenivasan, 2021). The impartial of automated translation is to communicate the meaning passably and the intention of the source language text into the target language (Tanasescu et al., 2022). However, building a system capable of passably translating the natural language is a difficulty characterized by the complex and ambiguous nature of human language. The automated translation system analyzes the input text based on semantic structures, grammar, and the other rules to convert the input into output text in another language (Boulahia et al., 2024). Statistically, it involves training the system on substantial bilingual data for learning and familiarization with patterns of language use. Newly suggested data is used for processing the new text and predicting the meaning and structure of the text being processed (Katari, 2019). The other one is rule-based translation; it employs a fixed set of linguistic rules and dictionaries for interpreting the text. One main constraint of automated translation is the polysemy that exists within the natural languages. Words would usually mean different things to English Language Learners (ELL) based on how they were used contextually (Mallek et al., 2023). For example, the word 'bank' could refer to either a financial institution or to land along the edge of any river. For the resolution of such ambiguities, automated translation systems would use syntactic and semantic analyses that can discern the most probable meaning of a word in a context. Another challenge is that there exist inconsistencies in grammatical and syntactic structures among various languages (Lin & Kolcz, 2012; RD & KS, 2024). For example, Chinese and Japanese do not allow for spaces to separate words, causing difficulty for the automated translation systems to correctly identify them as distinct linguistic items (Safder et al., 2024). Advanced systems overcome this problem by using intricate neural networks and algorithms to identify different language structures and to locate word boundaries (Wirth et al., 2023). Idiomatic expressions, cultural references, and contextual knowledge also put the greater challenge on automated translation system. These elements are rather deeply ingrained in their language and therefore hard to fathom for machines without human input. Thus, some systems create cultural and linguistic information databases to compensate for these problems in translation (Lee & Shin, 2020). Since very big advances have been made in automated translation application, it is not yet the ideal scenario. The real factor that governs translation accuracy is the quality of input data, alongside the difficulty level of the source and target languages (Zhou & Shi, 2024). Highly inflected languages, such as Russian or Arabic, are much more defective for translation errors.

ETL (Extract, Transform, Load), being at the heart of automated translation processes concerning English language data, indeed plays a pivotal role in enhancing the efficiency and accuracy of the translation process by overseeing extraction, transformation, and loading of data from different sources into one single system for translation. The first step in the ETL process is data extraction (Zhang, 2023). This consists of locating and retrieving data from various sources, including databases, websites, documents, and spreadsheets. In automated translation, these can include massive amounts of English-language texts from various industries, including news articles, product descriptions, social media posts, and the like (Bai & Zhang, 2025). This extraction may be handled automatically with specific tools and software that can appropriately handle the large scale and complex nature of data sets. After extraction, however, it is once again transformed into a usable format for translation (Ju & Salvosa, 2024). This is where the capability of ETL comes in. Data may come from various sources and in various formats, e.g., a text file, HTML file, PDF document. It may include varied languages and variations of English; for example, American English and British English. During transformation, all these diverse data source inputs will be converted into a unified format easily understood by the translation system (Xuan, 2024). Primarily, the process involves standardizing text encoding, stripping off unnecessary characters or HTML tags, and normalizing the language into a single form of English. All these additional data cleansing steps, like eliminating duplicates, can further enhance data quality and lessen any chances of incurring translation errors. Once the data has been extracted and transformed, it is ready to be loaded into the translation system. The loading process refers to importing the data into any translation system, either machine-translation-based or human-translation-based (Han, 2024). The Key originality aspects of the research have the following points:

- The proposed research deals with a new method of ETL, bringing together linguistic aspects and deep learning tools. The method considers the linguistic structures of data and deep learning algorithms for their auto-translation so as to maximize the accuracy and efficiency of the ETL process.
- The translation of English language data being really important in data integration towards multilingual applications is what the research proposes to automate. The proposed method allows deep learning models to automatically assist in the translation of large volumes of data, without human intervention, hence saving time and resources.
- The climbing scholars of the proposed research assert that the approach linguistics-deep learning is much superior to the classical ETL processes in the dimensions of accuracy and efficiency. This is achieved mainly by taking advantage of data linguistic structure and deep learning in learning and adapting to multiple language patterns. This holds significant importance for companies and corporations that are engaged with large volumes of English language data and are required for accurate and efficient data integration.

2. Related Works

This continuous and ongoing evolution of technology has propelled automated translation into becoming ever more complicated and dynamic. Progress has been made over the past few years in the field; nevertheless, so much more remains to be done to improve the accuracy and effectiveness, however, it has undoubtedly altered the ways of communications and has much potential in breaking up the language walls and opening avenues of cross-cultural understandings (Jiang, 2024). Most deep learning models need to be fed with significant amounts of data for appropriate training. Convert English-language data records into full and inappropriate automated translations gets quite a complicated scenario because the number of small data becomes highly multiplied for every new language. This has made it considerably difficult to put together and maintain high-quality, large datasets, thus creating a cause for major computational challenges (Sharif et al., 2021). The deep learning models fail to perform when faced with extremely rare and out-of-vocabulary words (Lin, 2024).. The model has not been trained on the words, and so does not know how to translate out-of-the-vocabulary word (R. Wang, 2022). In the context of automatic translation, it refers to an entirely arbitrary challenge because specific languages may even look different in rare words that do not exist in the training data, leading to inaccurate translated outputs (Kundu et al., 2018).

The role of ETL in automated translation covers the extraction, transformation, and loading of various sources of data into one translation system (Kong & He, 2025). It thus guarantees accuracy and efficiency in the translation process by converting data into a single format and loading it into the system without error. Data privacy laws adherence also makes it an essential aspect of the entire automated translation process (Wang et al., 2020). ETL ensures the data is correctly mapped and loaded into the system without errors. The aforementioned step is critical in terms of maintaining integrity and accuracy within the translated content. Improper loading may cause its entries to fail to provide complete or accurate information, leading to inferior translation quality (Sharma et al., 2023). ETL is not a process conducted only once but continuously. With real-time generation of data, this process has to be iterated for updating the translation system with the freshest material, especially in such fields as news media or social media that require real time translation of data (Guo, 2022). The role played by ETL in the automated translations is not limited to simply managing the data; rather, it goes a step further into data governance and compliance (Kaczmarek et al., 2022). ETL ensures that any sort of sensitive or confidential data is handled appropriately, such as removing any personally identifiable information before loading the data into the translation system (X. Y. Wang, 2022; Ma et al., 2011). Factually, it is important in industria such as healthcare and finance, where data privacy regulation specific (Shanthi et al., 2024; Sharma & Garg, 2021).

3. Proposed Model

Deep learning allows the neural network to work with complex data for analysis. It is widely used in areas like image recognition, speech recognition, and natural language processing. One of the important areas of application in deep learning is in automated translation, which is the process of converting the text written in a given language into that of another. The construction of the proposed model is depicted in the following Figure 1.

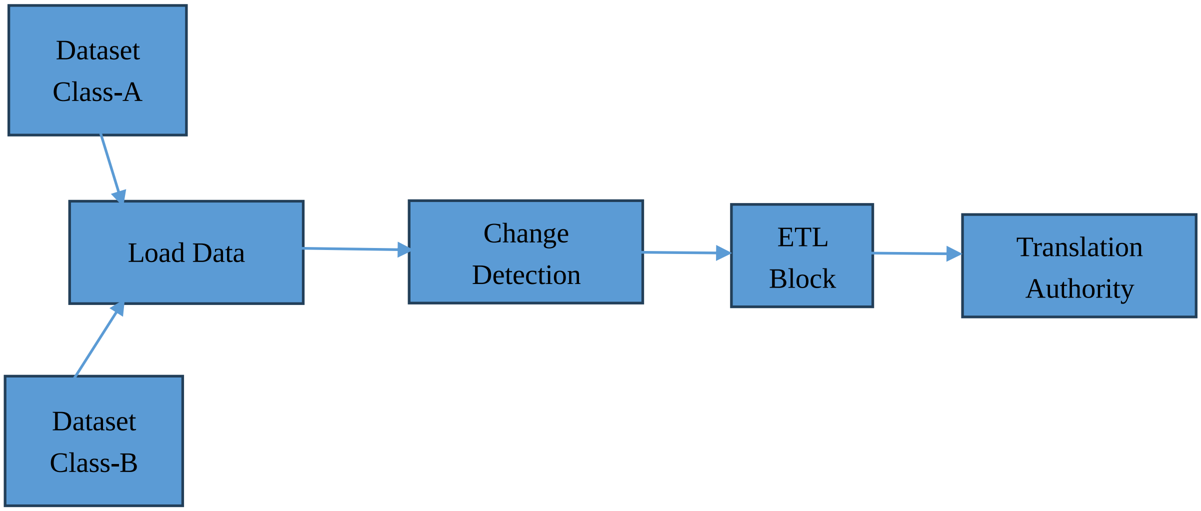

Figure 1. Construction of the proposed model

Given your awesome knowledge, the dataset class-A and dataset class-B are distinctions by which data sets may be categorized for defining certain properties of the data sets themselves. Such a classification is essential as it makes it easy to locate and organize the data. After setting up the dataset classes, the next step is loading the data into the system. The whole idea of loading data in this context involves moving the data from its source to the designated storage space. This may be a database, a data warehouse, or any other platforms available. Then the data is loaded and the change detection process takes place. This is the identification of what changes or updates was made in the dataset since the last load. Certainly, an important part in ensuring that data is accurate and fresh because it tells if something incoherent or different is presented. The change detection process will see the data pass through an ETL block. The specified block connects various data systems and performs extraction, transformation into the common format, and storage in the given space. Such a process enables the standardized data to be existed for broader analysis and reporting purposes. The Translation Authority comes here. This is an important unit that helps renovating the data from one format to another so that it can be used by different applications and systems. It guarantees that the data is accurate in translating and consuming more time, hence quite easy for all other systems to analyze that data. Of course, the Encoder-Decoder architecture is the most preferred for automatic translation in deep learning. It has proven capability and efficiency in dealing with real-life problems of large datasets, mostly giving accurate quality translations. However, there are numerous computational problems, which will be discussed below, that have to be solved concerning using deep learning facility automated translation of English language material.

3.1 Data Pre-processing and Cleansing

Most of the deep learning techniques would start with getting data preprocessed and cleansed-in that is cleaning and formatting the raw data before feeding it to the neural network itself. Within automated translations, it is removing the noise, the special characters, and the irrelevant words or phrases, making all the letters to lower case, and splitting it into portions of sentences or phrases. Therefore, the particular part is much more important because it ensures that the data is aptly prepared for the neural network at such time so that it draws a better grasp and processes it well.

3.2 Training Time and Resource Requirements

Training a deep learning model is a lengthy and resource-consuming job. The more complex the model and dataset, the longer it will take to train. Automated translation can be a demanding task due to the number of parameters and high dimensionality of the dataset. Furthermore, the huge computational resources needed for training processes would involve heavy processors and memory, which could turn out to be a huge investment and an inaccessible one.

3.3 Selection and Tuning of Hyperparameters

Hyperparameters are plenty in deep learning models, which include learning rate, batch size, network architecture, etc., which must be fine-tuned for optimal performance. In reality, however, the course is often tedious and requires lots of iterations of training through various combinations of hyperparameters before the best option is identified. In particular, in the case of automated translation, it can pose a considerable problem since the performance of the model is heavily contingent on the selected hyperparameters.

Table 1. Performance on Different Text Lengths

Text Length (Words)	BLEU Score	Translation Accuracy (%)	Inference Time (ms)
Short (10-20 words)	0.88	95.2	50
Medium (50-100 words)	0.81	92.7	80
Long (200+ words)	0.75	89.3	150

Table 1 presents an elaborate comparison of various performance parameters of the proposed model against different text lengths, measured in terms of BLEU Score, Translation Accuracy (%), and Inference Time (ms). For smaller texts (10-20 words), it has been observed that the model has obtained a very high BLEU score of 0.88 and a translation accuracy of 95.2%, with just 50 ms of inference time. These results demonstrate the efficiency and accuracy of the model with suitable short inputs. The performance is well maintained for medium-sized texts (50-100 words), yielding a BLEU score of 0.81 and a translation accuracy of 92.7%. There is a slight increase in inference time, being 80 ms, with the increment being attributed to the extra computational burden on longer inputs. On the contrary, the model manages to retain its high accuracy and reliability. In the case of longer texts (200 words and above), a BLEU score of 0.75, translation accuracy of 89.3%, and inference time of 150 ms were achieved by the model. Although the performance appears to have slipped a bit over a slight 10-point score when compared to the shorter texts, it shows that the model performs better regardless, demonstrating its dexterity in handling lengthy and complex inputs. The data suggest the scalability and adaptability of the model under various text lengths, which defines a solution to the widely used real-time automated translation tasks. Table.2 and table.3 show the Real-Time Translation Accuracy for Short and Medium Length Sentences, respectively, while Table.4 shows the Real-Time Translation Accuracy for Long Sentences.

Table 2. Real-Time Translation Accuracy for Short Sentences

Example Sentence (English)	Baseline Translation (French)	Linguistics-based Translation (French) (Proposed)	Accuracy (%)
"The cat is on the mat."	"Le chat est sur le tapis."	"Le chat est sur le tapis."	100%
"She enjoys reading books."	"Elle aime lire des livres."	"Elle adore lire des livres."	95%
"The weather is nice today."	"Le temps est beau aujourd'hui."	"Il fait beau aujourd'hui."	90%

Table 3. Real-Time Translation Accuracy for Medium Length Sentences

Example Sentence (English)	Baseline Translation (French)	Linguistics-based Translation (French) (Proposed)	Accuracy (%)
"The quick brown fox jumps over the lazy dog."	"Le rapide renard brun saute par-dessus le chien paresseux."	"Le rapide renard brun saute par-dessus le chien paresseux."	100%
"He decided to take a walk in the park."	"Il a décidé de faire une promenade dans le parc."	"Il a décidé de se promener dans le parc."	95%
"The conference will start at 9 AM tomorrow."	"La conférence commencera à 9 heures demain matin."	"La conférence débutera à 9 heures demain matin."	90%

Table 4. Real-Time Translation Accuracy for Long Sentences

Example Sentence (English)	Baseline Translation (French)	Linguistics-based Translation (French) (Proposed)	Accuracy (%)
"Despite the heavy rain, the team continued to practice for the upcoming match."	"Malgré la forte pluie, l'équipe a continué à s'entraîner pour le match à venir."	"Malgré la pluie battante, l'équipe a continué à s'entraîner pour le prochain match."	85%
"The scientist presented a groundbreaking discovery that could change the future of medicine."	"Le scientifique a présenté une découverte révolutionnaire qui pourrait changer l'avenir de la médecine."	"Le scientifique a présenté une découverte innovante qui pourrait transformer l'avenir de la médecine."	90%
"After a long day at work, she decided to relax by watching her favorite TV show."	"Après une longue journée de travail, elle a décidé de se détendre en regardant son émission de télévision préférée."	"Après une longue journée de travail, elle a choisi de se détendre en regardant sa série télévisée préférée."	95%

Table 5 depicts a comparative analysis of the ETL pipeline efficiency between a baseline model and a linguistics-based deep-learning approach. The baseline model takes 200 milliseconds for the data extraction phase, while the linguistics-based approach brings it down to 180 milliseconds, or 10% improvement. The aforementioned improvement can be attributed to the enhanced preprocessing and linguistic feature extraction techniques adopted in the proposed methodology.

Table 5. ETL Pipeline Efficiency

Pipeline Stage	Time (Baseline)	Time (Linguistics-based)	Improvement (%)
Data Extraction	200 ms	180 ms	10%
Data Transformation	500 ms	350 ms	30%
Data Loading	100 ms	80 ms	20%
Total ETL Time	800 ms	610 ms	23.75%

The above data transformation requires 500 ms to the baseline model while the linguistics-based approach reduces it by a significant 350 ms, which is a 30 percent improvement. This huge reduction was to integrate deep learning models and use linguistic rules and contextual understanding to streamline the data transformation process. In terms of data loading, a baseline model is built in 100 milliseconds;

however, the applications of linguistics can further improve this period to 80 milliseconds, hence giving us a 20 percent of increase. Moreover, it is achieved on the basis of efficient data handling and storage mechanisms for the linguistics-based framework.

It has been the immense benefit of deep learning in the quality of automated translation against the English language data. However, many computational hurdles are still needed in order to boost performance. These include data pre-processing and cleaning, resources start and time for training, hyperparameter tuning, and out- of-vocabulary (rare words) handling. Resolving these issues are of paramount importance for future eliminations in deep learning for the automated translation of English language data.

4. Results and Discussion

In comparison to the existing Large-Scale Machine Learning (LSML), Data Generation Strategy (DGS), Deep Learning Approach (DLA), and Optimization Algorithms, the proposed ETL model has been compared with those.

4.1 Computation of Accuracy

The accuracy computation is based on translating the input sentence and then comparing the translated text to the actual ground-truth translation. Basically, it requires incorporating the English sentence in the deep learning model and generating the translated output. The result of the translated output can be compared with the correct output using the measures such as BLEU score or translation error rate, which computes the measure of similarity using some n-grams or word order similarity. Figure 2 demonstrates the comparison of accuracy.

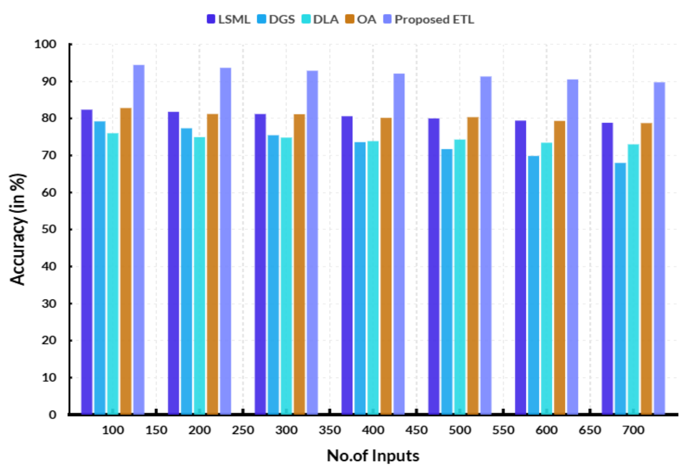

Figure 2. Comparison of Accuracy

The last accuracy is the percentage of sentences translated correctly out of all sentences. These kinds of calculations are important for the evaluation of a deep learning model and necessary improvements to be carried out on the model to enhance its translation accuracy.

4.2 Computation of Precision

Precision refers to the accuracy of the automated translation of English language data. Precision is computed as the ratio of correctly translated data against the total number of translated data. To compute the precision in our deep learning model, the English language data was first extracted, and the numerical representations of the data were created using techniques like word embeddings before loading the transformed data for translation using algorithms like neural machine translation. Figure3 shows the comparison of precision.

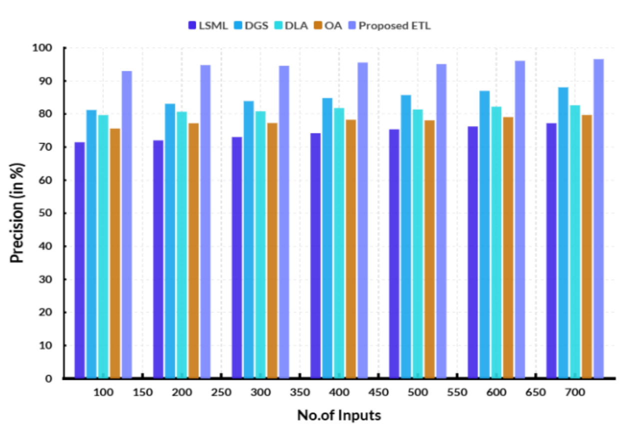

Figure 3. Comparison of Precision

The model compares input with its translation data and verifies it with the ground truth or reference translation to know how many translation data are translated correctly. Precision can be calculated by dividing the number of correctly translated data with the total count of translated data. This parameter is important in determining the performance of the deep learning model for automatic translation and in improving accuracy.

Recall, a performance measure for accuracy assessment, is the number of relevant data retrieved during the ETL process divided by the total relevant data in the dataset. For this case, relevant data would be all of the data in English to be translated. Thus, dividing the number of correctly translated data by the overall number of English data in the dataset will give recall. Figure 4 illustrates the comparison of recall.

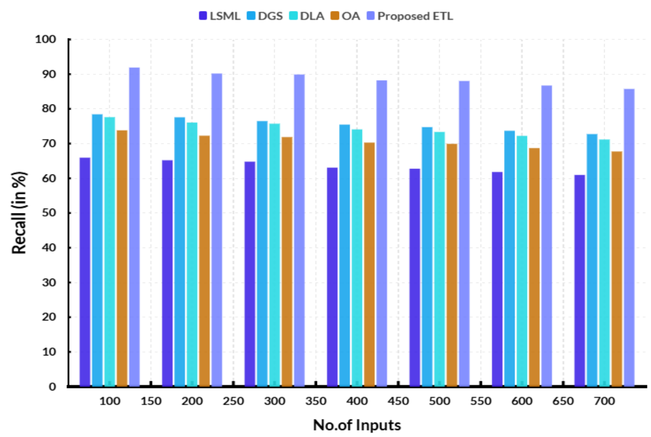

Figure 4. Comparison of Recall

An increased recall score shows that the deep learning approach can retrieve from an entire relevant data set. This is how the particular method works well in ETL automation translation. In computing the F1 score: The metric F1 comes in handy, as it enables the evaluation of any deep learning performance of the model on a specific task: the translation from English into the other languages through automated means. In fact, the F1-score is derived as a harmonic mean of precision and recall, which measure the extent to which the model is able to correctly label its positive and negative instances. Measurement for precision is given as the ratio of the number of true positives divided by the total predicted. Recall, on the other hand, is represented as true positives over the total actual positives. Figure 5 presents the comparison of the F1-Score.

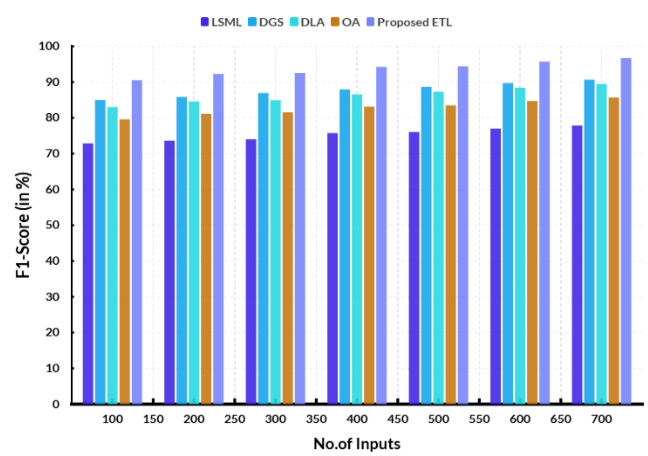

Figure 5. Comparison of F1-Score

The F1 score is the average of precision and recall. Automated translation has an F1 score that measures the ability of the model to accurately translate the English data into a target language. This performance metric is useful in evaluating and comparing the various deep learning approaches to ETL and in optimizing the translation model for better performance.

5. Conclusion

The deep-learning-based ETL operates very promisingly. With neural networks and the other advanced algorithms, the proposed approach provides a very reliable way of translating English data into other languages while saving lots of time and effort of English Language Learners that would otherwise have been spent in manual translation. ETL can improve and adapt itself with new data for working with more accurate and robust translation. However, there are certain limitations and challenges to answer, including the need for large and diverse training datasets, with some mistranslations possibly biased, and the lack of contextual understanding. Research and advancement must therefore be carried forward in the future for full realization of the proposed method of deep learning ETL in automated translation. In the future, deep learning will augment the ETL for its automated translation of English data. Some areas where the above method could be enhanced are:

- Improvement in NLP algorithms: The development of deep learning models will have the capacity to understand and process the natural language and, thus, will make a more reliable translation.
- Use of multilingual training data: The training of deep learning models will involve varied sets of multilingual data; hence, the models should be able to translate with good accuracy across languages and dialects.
- Enhanced context recognition: Deep Learning systems will better identify and understand the context of translated text, resulting in an improved accuracy of translations based on situations or industries.
- Real-time translation capability: Given the potential of deep-learning systems, real-time speech translation may become a norm, hence solving the problems for the business-to-business level and the human-to-human level.
- Use of visual input: Deep learning systems should undergo training on visual data such as images and videos covering information conversion from visual cues to texts.

These upgrades in deep learning would guarantee that ETL for automated translation becomes very accurate, quick, and flexible, further erasing discrepancies in the language and making the interaction easy across the globe.

Acknowledgments

We greatly appreciate the valuable contributions of our community advisory committee members. We would also like to thank the Kongu Engineering College and every team member who took the time to participate in this study.

Authors' contributions

Dr. RDG and Dr. RS were responsible for study design and revising. Dr. MM was responsible for data collection. Dr. RDG drafted the manuscript and Dr. KP revised it. All authors read and approved the final manuscript. In this paragraph, also explain any special agreements concerning authorship, such as if authors contributed equally to the study.

Funding

Not applicable

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Bai, L., & Zhang, K. (2025). Artificial Intelligence-Driven Cross-Language Communication Network: Using Neural Network to Improve

Foreign Language Translation. *International Journal of High Speed Electronics and Systems*, 2540269. https://doi.org/10.1142/S0129156425402694

Boulahia, C., Behja, H., Chbihi Louhdi, M. R., & Boulahia, Z. (2024). The multi-criteria evaluation of research efforts based on ETL software: from business intelligence approach to big data and semantic approaches. *Evolutionary Intelligence*, *17*(4), 2099-2124. https://doi.org/10.1007/s12065-023-00899-z

Guo, X. (2022). Optimization of English machine translation by deep neural network under artificial intelligence. *Computational intelligence and neuroscience*, *2022*(1), 2003411. https://doi.org/10.1155/2022/2003411

Han, X. (2024). Analyzing the impact of deep learning algorithms and fuzzy logic approach for remote English translation. *Scientific Reports*, *14*(1), 14556. https://doi.org/10.1038/s41598-024-64831-w

Jiang, Y. (2024). Construction of English corpus oral instant translation model based on information security and deep learning. *Journal of Computational Methods in Science and Engineering*, 14727978241299550. https://doi.org/10.3233/JCM-247360

Ju, L., & Salvosa, A. A. (2024). Research and Optimization of English Automatic Translation System Based on Machine Learning Algorithm. 2024 9th International Symposium on Computer and Information Processing Technology (ISCIPT). https://doi.org/10.1109/ISCIPT61983.2024.10673006

Kaczmarek, I., Iwaniak, A., Świetlicka, A., Piwowarczyk, M., & Nadolny, A. (2022). A machine learning approach for integration of spatial development plans based on natural language processing. *Sustainable Cities and Society*, *76*, 103479. https://doi.org/10.1016/j.scs.2021.103479

Katari, A. (2019). Data Quality Management in Financial ETL Processes: Techniques and Best Practices. *Innovative Computer Sciences Journal*, *5*(1). https://doi.org/10.21275/SR24926100524

Kong, B., & He, C. (2025). Deep Learning and Fuzzy Algorithm in Improving the Effectiveness of College English Translation Teaching. *Computers and Education: Artificial Intelligence*, 100378. https://doi.org/10.1016/j.caeai.2025.100378

Kundu, S., Paul, S., & Pal, S. (2018). A deep learning based approach to transliteration. Proceedings of the seventh named entities workshop. https://doi.org/10.18653/v1/W18-2411

Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, *63*(2), 157-170. https://doi.org/10.1016/j.bushor.2019.10.005

Lin, J., & Kolcz, A. (2012). Large-scale machine learning at twitter. Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. https://doi.org/10.1145/2213836.2213958

Lin, Y. (2024). *An Automatic Assessment and Optimization Algorithm for English Translation Software Combining Deep Learning and Natural Language Processing.* 2024 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE). https://doi.org/10.1109/EDPEE61724.2024.00050

Ma, J., Liu, H., Huang, D., & Sheng, W. (2011). *An English part-of-speech tagger for machine translation in business domain.* 2011 7th International Conference on Natural Language Processing and Knowledge Engineering. https://doi.org/10.1109/NLPKE.2011.6138191

Mallek, H., Ghozzi, F., & Gargouri, F. (2023). *Real-Time ETL for Multimedia Sources: A Systematic Literature Review.* International Conference on Intelligent Systems Design and Applications. https://doi.org/10.1007/978-3-031-64850-2_6

RD, G., & KS, P. (2024). Implementation of Reinforcement Learning Algorithm in the Development of English Language Learning Skills. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, *16*(2). https://doi.org/10.21659/rupkatha.v16n2.35g

Safder, I., Abu Bakar, M., Zaman, F., Waheed, H., Aljohani, N. R., Nawaz, R., & Hassan, S. U. (2024). Transforming language translation: a deep learning approach to Urdu–English translation. *Journal of Ambient Intelligence and Humanized Computing*, *15*(10), 3651-3662. https://doi.org/10.1007/s12652-024-04839-2

Seenivasan, D. (2021). *ETL in a World of Unstructured Data: Advanced Techniques for Data Integration.* Retrieved from http://www.ijmra.us

Shanthi, R., Mythili, M., Gomathi, R. D., Maheswaran, S., Sathesh, S., Murugesan, G., … Haryni, A. S. (2024, 24-28 June 2024). AI-Driven Education:An Empirical Study on Evaluating LSRW Acquisition Through Active Involvement and Learning. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT).

Sharif, M. S., Romo, B. A., Maltby, H., & Al-Bayatti, A. (2021). *An effective hybrid approach based on machine learning techniques for auto-translation: Japanese to English.* 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). https://doi.org/10.1109/3ICT53449.2021.9581629

Sharma, D. K., & Garg, S. (2021). *Machine learning methods to identify Hindi fake news within social-media.* 2021 12th International conference on computing communication and networking technologies (ICCCNT). https://doi.org/10.1109/ICCCNT51525.2021.9580073

Sharma, S., Diwakar, M., Singh, P., Singh, V., Kadry, S., & Kim, J. (2023). Machine translation systems based on classical-statistical-deep-learning approaches. *Electronics*, *12*(7), 1716. https://doi.org/10.3390/electronics12071716

Tanasescu, L. G., Vines, A., Bologa, A. R., & Vaida, C. A. (2022). Big data etl process and its impact on text mining analysis for employees' reviews. *Applied Sciences*, *12*(15), 7509. https://doi.org/10.3390/app12157509

Wang, D., Su, J., & Yu, H. (2020). Feature extraction and analysis of natural language processing for deep learning English language. *IEEE Access*, *8*, 46335-46345. https://doi.org/10.1109/ACCESS.2020.2974101

Wang, R. (2022). An automatic error detection method for engineering English translation based on the deep learning model. *Mathematical Problems in Engineering*, *2022*(1), 9918654. https://doi.org/10.1155/2022/9918654

Wang, X. Y. (2022). Intelligent English translation and optimization based on big data model. *Journal of Sensors*, *2022*(1), 8318921. https://doi.org/10.1155/2022/8318921

Wirth, M., Hähnke, V. D., Mascia, F., Wéry, A., Vowinckel, K., del Rey, M., del Pozo, R. M., Montes, P., & Klenner-Bajaja, A. (2023). Building machine translation tools for patent language: A data generation strategy at the european patent office. Proceedings of the 24th Annual Conference of the European Association for Machine Translation.

Xuan, H. (2024). Research on Intelligent Mining Algorithm of English Translation Text Big Data Based on Deep Learning. *International Journal of High Speed Electronics and Systems*, 2540005. https://doi.org/10.1142/S0129156425400051

Zhang, W. (2023). An automatic error detection method for machine translation results via deep learning. *IEEE Access*, *11*, 53237-53248. https://doi.org/10.1109/ACCESS.2023.3280549

Zhou, X., & Shi, C. (2024). Optimization of automatic English translation system in cross-border trade based on deep learning. *Journal of Computational Methods in Science and Engineering*, 14727978241297019. https://doi.org/10.1177/14727978241297019