



**Faculty of Education**

**Journal of Education**

\*\*\*

**The Uses of Machine Learning (ML) in Teaching  
and Learning English Language: A Methodical  
Review.**

**BY**

***Dr. Abdullah AbdulMuhsen AlHarbi***

**Receipt date: 3 November 2021 - Date of acceptance: 28 November 2021**

**DOI: 10.12816/EDUSOHAG. 2021.**

## **Abstract**

**Regular teaching methods for ESL/EFL may rely on assumptions that do not lead to the desired results in student learning. Consequently, the current study sought to perform a systematic review of the research that revealed numerous applications of machine learning (ML) in English language instruction (Applied Linguistics) and the development of interpretable machine learning or associated practical software. The current research employed a mixed-methods approach (approach) that included qualitative and quantitative analysis to survey the previous studies on machine learning in EFL/ESL instruction. Results showed that machine learning (ML) may teach English as a foreign language or as a second language and that it has a bright future in this sector. Interpretable machine learning (ML) was found to help in teaching many EFL/ESL topics such as dictation, phonetics instruction, language blogging, and all of which contribute to the facilitation of English language learning and teaching processes. Finally, the study concluded with recommendations for researchers in EFL or ESL on how to incorporate machine learning in English classrooms using functional programming languages, tools, and platforms. More studies were also advised to determine the most effective methods and procedures for incorporating machine learning (ML) into English as ESL/EFL in the classroom.**

**Keywords: artificial intelligence- teaching ESL/ EFL- teaching strategies- applied linguistics- phonetics.**

## استخدامات التعلم الآلي (ML) في تدريس وتعلم اللغة الإنجليزية: مراجعة منهجية

### المستخلص

الأساليب التقليدية في تدريس اللغة الإنجليزية قد تعتمد على افتراضات لا تؤدي إلى تحقيق النتائج المطلوبة في تعلم الطلاب. وهذا الأمر جعل هناك حاجة للبحث عن طرق أكثر تقدمًا للتغلب على أوجه القصور في تدريس اللغة المعقدة التي نشأت وتتنشأ في المواقف المختلفة لتدريسها وتعلمها. لذلك، هدفت الدراسة الحالية إلى إجراء مراجعة منهجية للبحوث التي كشفت عن كثير من استخدامات تطبيقات التعلم الآلي (ML) في تدريس اللغة الإنجليزية. وقد استخدم البحث الحالي منهجا يخلط بين البحث النوعي والكمي لمسح الدراسات السابقة حول التعلم الآلي في تدريس اللغة الإنجليزية كلفة أجنبية / كلفة ثانية. وأشارت النتائج إلى دور استخدام التعلم الآلي (ML) في تحسين تعلم اللغة الإنجليزية كلفة أجنبية / كلفة ثانية. وكشفت النتائج أيضًا عن إمكانية استخدام التعلم الآلي (ML) لفصول اللغة الإنجليزية كلفة أجنبية / كلفة ثانية في تدريس الإملاء، والصوتيات، وعمل المدونات اللغوية، وكثير مما يدعم تسهيل عمليات تدريس وتعلم اللغة الإنجليزية. أخيرًا، ختمت الدراسة بتوصيات للباحثين في تدريس اللغة الإنجليزية كلفة أجنبية (EFL) أو كلفة ثانية (ESL) لاستخدام لغات وأدوات ومنصات البرمجة الوظيفية لتنفيذ التعلم الآلي في فصول اللغة الإنجليزية. كما أوصت بإجراء مزيد من الدراسات للبحث عن أفضل الطرق والأساليب والإجراءات لاستخدام التعلم الآلي (ML) في تدريس اللغة الإنجليزية كلفة أجنبية / اللغة الإنجليزية كلفة ثانية.

الكلمات المفتاحية: الذكاء الصناعي - تدريس اللغة الإنجليزية - استراتيجيات التدريس - اللغويات التطبيقية - الصوتيات.

**Introduction:**

Global connection, intelligent machines, and new media are a few factors driving the 4th Industrial Revolution. This symbolizes a new phase in the organization and control of the industrial value chain (Industry 4.0). Because of the pervasiveness of the Internet during the fourth industrial revolution, society has evolved to be more social and more knowledgeable, and the globe has become smaller than it was previously. Educational reform thus serves to meet the needs of Industry 4.0, in which humans and machines work together to generate new possibilities in the field of education. Learning skills in education must combine life skills and innovative skills and a combination of 21st-century skills, including leadership, collaboration, creativity, digital literacy, productive communication, emotional intelligence, entrepreneurship, global citizenship, problem-solving, and teamwork, among other things. Learners are more likely to use intelligent agents, mobile technologies, and cloud computing to promote interactive education if they combine these abilities in their arsenal.

The education industry has incorporated artificial intelligence as an educational tool utilized inside and outside the classroom with technological advancements. Using artificial intelligence creates an interactive learning environment. Therefore, this technology can make the learning experience more student-centered, meaningful, and capable of encouraging students to explore their creativity while encouraging innovation for both teachers and students.

In recent years, the value of statistical approaches in EFL/ESL teaching and learning has been increasingly recognized (Lin, 2021). This could be that statistical methodologies are crucial in researching factors in EFL/ESL teaching and learning. One of the largely often employed approaches is linear regression. However, the approach is hampered by solid assumptions like normality, linearity, and homoscedasticity (Plonsky & Ghanbar, 2018). As a result, increasingly powerful algorithms are being used to address more complex EFL/ESL teaching and learning difficulties. Machine learning represents these new technologies to a considerable extent (ML). On the other hand, there are no systematic reviews of ML in EFL/ESL teaching and learning. Most studies on the methodologies used in EFL/ESL teaching and learning focus on classic methods, such as linear regression.

Furthermore, most EFL/ESL researchers are unsure of how to report and understand the results of ML models. It is necessary to introduce corresponding methods for utilizing machine learning. As a result, this review pursues to fill these gaps by outlining the uses of machine learning in EFL/ESL teaching and learning, proposing interpretable machine learning, and presenting suggested methodologies for researchers in EFL/ESL teaching and learning.

### **Statement of the research problem:**

Modern trends resulting from the various changes that affected learning and education after openness to technology, interaction with social media platforms, and ease of communication with the outside world, contributed to expanding students' perceptions, keeping pace with the skills of the twenty-first century. These technologies became able to transform education from banking education, Squishy to open, investigative skill education.

As well as the difficulties that the researcher experienced in teaching and learning the English language at the time of his teaching, it made an urgent need to think of new methods to be used in facilitating the processes of teaching and learning the English language and evaluating the progress of learners in it.

Some studies, such as (Settles et al., 2020), also recommended using modern technologies, especially machine learning technology. It can benefit from relying on its programming on logarithms in storing many of the English language vocabulary, grammar, and skills. Settles et al. (2020) reported that ML may formulate teaching models to enhance students' learning and evaluate their progress in learning English.

### **Questions of the study:**

The problem of the current research can be represented in the following main question:

What are the uses of machine learning in teaching EFL/ESL?

The following sub-questions emerge from this question:

2. How ML can treat the interpretability in EFL/ESL teaching?

3. What is the appropriate orientation for EFL/ESL academics to benefit from machine learning?

**The objectives of the research:**

The objectives of the current research are:

1. Studying the efforts of ML in teaching EFL / ESL.
2. Recognizing the linguistic contributions made by machine learning in the teaching of ESL/EFL.
3. Determining the interpretability of ML in Teaching ESL/EFL.
4. Discovering the proper orientation for EFL/ESL academics to take advantage of ML.

**The importance of the study:**

A significant study subject in artificial intelligence, machine learning can effectively evaluate and deeply mine data from teachers and students in the education industry application, opening new development opportunities for EFL/ESL teaching and learning in the future and other fields. According to the researcher's findings, researchers in English language education and learning have only rarely investigated the subject matter of this study. A guide for future relevant efforts in applying machine learning in EFL/ESL teaching and learning is intended to be developed from this research.

**Literature Review:****Definition of ML:**

As Hurwitz and Kirsch (2018) defined, machine learning is a subset of artificial intelligence that helps a computer system to learn from data rather than explicit programming. On the other hand, machine learning is a possible but not an easy process. Machine learning can be cataloged into two types: supervised and unsupervised learning. Alruily (2021) classified text processing machine learning into three fundamental subcategories: supervised, unsupervised, and semi-supervised. Supervised learning is the most common type of machine learning (p.10).

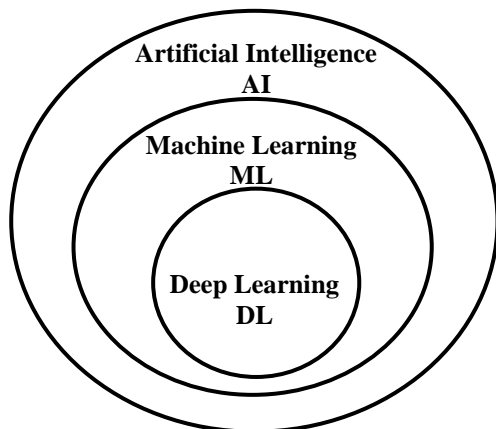


Figure 1: ML as a part of AI (Andoni, 2020)

According to Wang (2019) machine learning, for example, can assess data from students’ learning situations and successes over the past year and generate related written reports. Based on this vital data, teachers and students can examine the difficulties and causes of learning and then provide the necessary learning programs and promotion strategies. It was stated by Liu (2021) that the application of science and technology project learning in the English linguistics curriculum, together with the integration of machine learning, resulted in a significant increase in the learners' level of proficiency in applied linguistics.

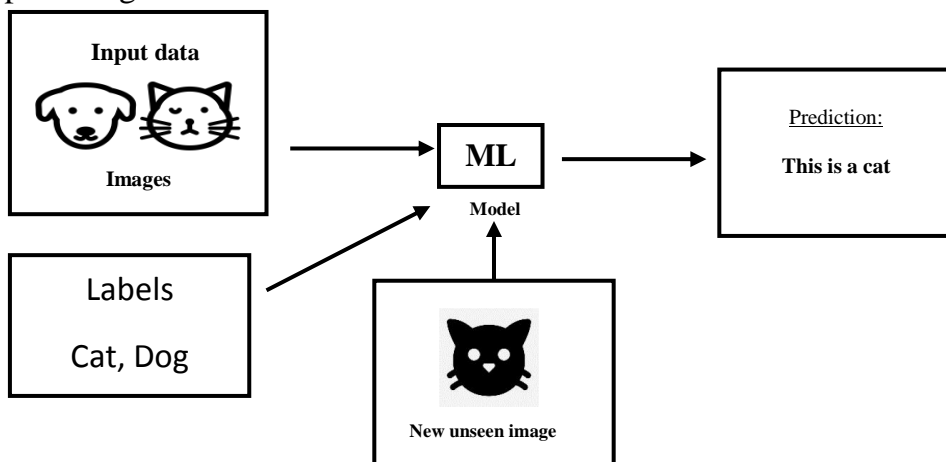


Figure 2: supervised learning (Loukas, 2020)

The importance of algorithms in machine learning was also discussed by Hurwitz and Kirsch (2018). They made sure that no discussion of machine learning was complete without mentioning algorithms. The most common programming languages for machine learning algorithms are Java, Python, and R. Machine learning libraries are available in each of these languages, which enable a wide range of machine learning algorithms. Furthermore, these languages have active user communities that regularly contribute code and debate ideas, issues, and solutions to educational problems. Li (2020) conducted a quasi-experimental study to determine the usefulness of ML in the teaching of applied linguistics at the university level. The findings revealed that the linguistic grades of the experimental class were significantly higher than those of the control class in the study.

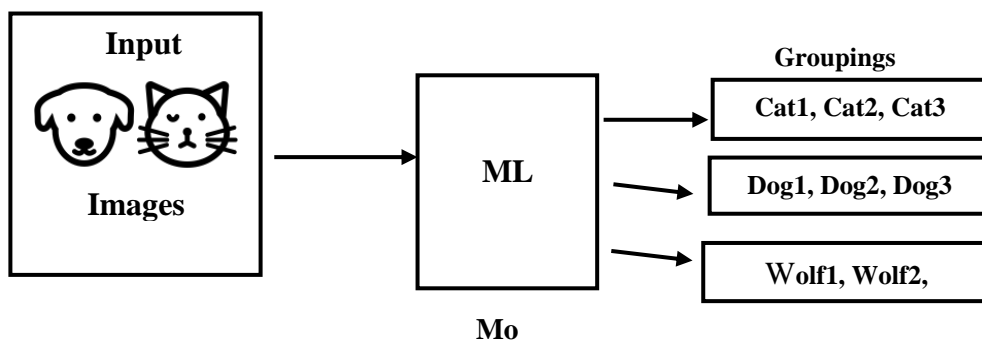


Figure 3: unsupervised learning (Loukas, 2020)

### Previous studies on old methods of teaching ESL/EFL:

Plonsky and Oswald (2017) conducted a thorough evaluation of manifold regressions in ESL (L2) research, and they compared their findings to those obtained by ANOVA. Specifically, Meteyard and Davies (2020) studied linear mix-effects models to understand them better. They expressed their reservations about using this strategy in psycholinguistics. Outliers are a challenge in L2 research, according to Nicklin and Plonsky (2020), who outlined the problem and summarized the current strategy for dealing with outliers. Plonsky and Gonulal (2015) conducted an appraisal of the investigative element of exploration in



studying languages. Li and Stromberg (2016) examined inferential statistics in ESL/EFL education research and concluded that p values are inappropriate. Norris (2015) reviewed statistical significance testing in L2 research, identifying the difficulty with statistical testing, and advocating for reform. The development trend of the research approach used in L2 research was summarized by King and Mackey (2016). The quantum approaches in corpus linguistics, comprising ANOVA, element analysis, and resampling, were reviewed by Paquot and Plonsky (2017).

Furthermore, scholars such as Norouzian (2020) studied sample size planning in ESL research. To cope with non-normal data and outliers, Nikitina and Furuoka (2018) suggested using quantitative regression with bootstrapping. These reviews and studies made significant additions to our grasp of applied linguistics methodology. However, the issue is that this research is solely focused on traditional approaches. The tight presuppositions have not been overcome, and the desire for more advanced applied linguistics teaching and learning procedures to deal with more intricate issues has not been met. This frees up space for research into machine learning in applied linguistics.

### **Research Method:**

The current study employed one of the mixed-methods research strategies, which is (Concurrent Embedded Design). This strategy is based on collecting data using quantitative and qualitative methods concurrently in a single stage, with a primary focus on gathering scientific research data and providing the basic method or methods more importance and weight. It is not essential as the other or secondary approach, which is done simultaneously as the preceding method and may be used to answer a new research issue or conduct another partial study within the study.

### **procedures of the study data:**

To accomplish the objectives of the research and answer its question, the researcher applied the following procedures:

### **1. Standards of Inclusion:**

The following are the inclusion requirements that must be met: First and foremost, machine learning should solve significant problems. Second, the challenge should have something to do with the process of acquiring applied linguistics techniques. Third, studies that can give a spotlight on how to employ machine learning to asynchronous learning are also included.

### **2. Previous Research:**

The data are gathered utilizing the Saudi Digital Library (SDL), Scimago Journal & Country Rank (SJR), Web of Science, and Google Scholar. To begin, the researcher conducts a keyword search for terms relating to machine learning, teaching artificial intelligence, and artificial intelligence. Additionally, exemplary research from related disciplines is included, but they represent a minor part of the total. The inclusion of these papers demonstrates how to do multi-disciplinary studies involving ESL/EFL learning and machine learning.

### **3. Standards of Exclusion:**

Due to the research's focus on the uses of ML in teaching ESL/EFL, the research should be connected to both machine learning and ESL/EFL teaching. If the study focuses exclusively on one element, it will be discarded.

### **4. Coding and grouping data of the research:**

Many studies were reviewed by the researcher one by one after being investigated and dismissed several times. Because of the large number of distinct fields within ESL/EFL, the results are grouped based on how similar they are. The reviewed studies were analyzed thematically to five themes: ESL research, language evaluation, ESL/EFL Teaching & learning, Corpus linguistics, and Phonetics learning.

Discussion of the results:

To respond to the study's first question, Table 1 presents a descriptive analysis of the subjects identified through an objective investigation of the target studies: ESL research, language assessment, ESL teaching and

learning, linguistics and phonetics learning, Corpus linguistics, and phonetics learning. In the studies reviewed for this study.

Table 1  
descriptive analysis of reviewed topics

Topic	Frequency	Percent
ESL research	4	12.1
Language evaluation	8	24.2
ESL/EFL Teaching & learning	7	21.2
Corpus linguistics	9	27.3
Phonetics learning	5	15.2
Total	33	100.0

Table (2)

indicates a description of the methodologies that reviewed research employed.

Table 2: descriptive analysis of research methodologies.

Method	Frequency	Percent
quantitative	3	9.1
qualitative	11	33.3
Mixed-method	11	33.3
quasi-experiment	8	24.2
Total	33	100.0

**ESL/EFL Research:**

The Bayesian analysis algorithm may be the most talked-about machine learning algorithm in second language studies (L2). Gudmestad and colleagues (2013) explained how Bayesian analysis might be employed in second language attainment.

Norouzian and colleagues (2018) investigation is comparable to the one carried out by Gudmestad and colleagues. It was discovered in their research that Bayesian networks might be a very effective instrument in L2 research, and they discussed that the beginning of a Bayesian revolution in second language research had begun. Clustering, in addition to Bayesian networks, is extremely useful in the study of L2 acquisition. Papi and Teimouri (2014) classified the motivations of language learners by employing clustering techniques. It has been demonstrated by Warschauer et al. (2019) that gathering algorithms can

aid in the learning of new vocabulary terms. Crowther and colleagues (2020) undertook a thorough review of the applications of clustering algorithms in second-language studies, published in the second Language Studies journal. According to the facts offered above, Bayesian analysis has been used in L2 study by linguists to make decisions. Student clustering can assist us in dividing students into various groups depending on their needs, which can then be used to personalize language education and learning to meet their specific requirements.

**Language learning evaluation:**

Machine learning applications in language testing can be separated into three categories: test administration, automatic essay scoring, and cognitive diagnostic evaluation. Language testing is a large field, and machine learning applications in language testing can be divided into three areas. According to Chapelle and Chung (2010), who reviewed all the options, normal language processing and speech recognition have been utilized in language assessment.

Rudner (2016) investigated Bayesian probability theory in the context of test administration and discovered that it might be beneficial to give feedback to students irrespective of the exam's size. When it comes to robotic marking, Latifi and Gierl (2020) employed machine learning and everyday language treating algorithms to estimate the score based on Coh-Metrix attributes automatically.

Convolutional neural networks were used to improve the essay grading system, according to Shin and Gierl (2020), and the results showed that neural networks outperformed traditional approaches. Thomasevic and colleagues (2020) looked at how supervised machine learning may be applied to expect student exam performance. Litman et al. (2018) examined the methods, problems, and opportunities that current technology, notably machine learning-based voice recognition, has brought to the assessment of the second language speaking.



*Figure 4: An example for dictation test using ML technology (Settles et al., 2020, p.252)*

Zheng et al. (2020) used machine learning techniques to create an adaptive test system. Almond et al. (2007) illustrated how Bayesian networks might be used to simulate CDA utilizing Bayesian networks in the context of cognitive diagnostic evaluation (CDA). Cui et al. (2016) employed a neural network in small sample sizes. Automatic essay marking and robotic speech grading, both depending on machine learning tools, is anticipated to be an innovative newborn tendency in the future, and learner capability may be identified more exactly as machine learning technology increases in sophistication.

### **Language teaching and language learning:**

Machine learning's impact on language teaching and learning is visible in technology-enhanced teaching and learning. Crossley (2013) conducted research that meticulously depicted computational tools and machine learning to investigate second language writing. Wiechmann and Kerz (2014) investigated the possibility of using squad learning in German works. It has been established that resolution trees and the random woodland algorithm can assist us in comprehending what we are reading.

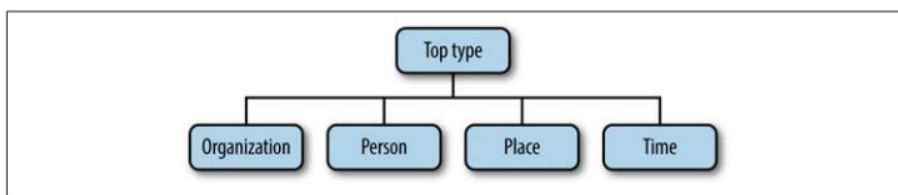
Kelly et al. (2018) focused on real-world classroom issues. He used machine learning to develop a model that can automatically assess the legitimacy of classroom questions. Yang and Li (2018) utilized backpropagation neural networks to predict learner accomplishment and discover elements influencing their outcome. Pliakos et al. (2019)

suggested a solution to the adaptive learning cold start problem by merging machine learning with item response theory.

Rico-Juan et al. (2019) investigated peer evaluation in the classroom using machine learning and discovered a mismatch between arithmetic marks and reaction. Charitopoulos et al. (2020) described the usages of information reproducing techniques, demonstrating how these algorithms could be used to improve language instruction in the future. Even though some examples are from nearby, they serve as good language teaching and learning models. These examples can be used to guide future research on machine learning and language training. Finally, machine learning-enhanced systems may improve language teaching and learning.

### **Corpus Linguistics:**

Correspondingly, corpus linguistics is a methodology that uses large, readily available collections of naturally occurring spoken and written texts (known as corpora) to perform computer-based empirical investigations of language use (both quantitative and qualitative). The random forest approach discussed in greater detail below is perhaps the most extensively used algorithm in corpus linguistics. Th Gries completed his Ph.D. by systematically reviewing the random forest model on corpus linguistics data (2020). Fonteyn and Nini (2020) utilized the random forest technique and restrictive inference trees, both of which they built themselves, to analyze gerunds in their research. Deshors (2020a) used random forests to analyze the contextualized past tense and the exchanges between variables in his study. Using this strategy, Deshors argued that it is possible to override the supposition of normality, which had previously been addressed in the previous studies review, and that it is more accurate. Using random forests to evaluate the interactions of numerous speakers in 2020, he did yet another investigation on the topic.



*Figure 5: An example for syntactic tree structure using ML technology (Pustejovsky& Stubbs, 2013, p.19)*

Moreover, using support vector machines, Sung et al. (2015) were able to categorize the legibility of L2 reading texts concerning the Common European Framework of Reference for Languages (CEFR). Using ML techniques, H. Kang and Yang (2020) were able to quantify the political bias present in their findings. According to Ballier and colleagues (2020), a Kaggle competition used ML algorithms and natural language processing procedures to rate essays robotically, and the findings were published in the journal Science. According to all available evidence, the haphazard forest approach is the most often exploited algorithm in corpus linguistics, and that other machine learning algorithms have substantial potential.

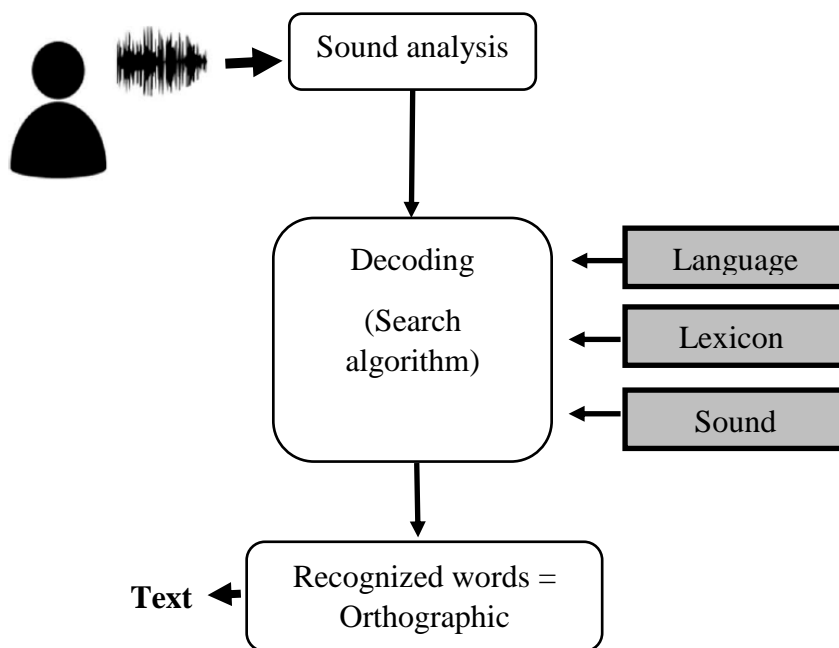
Table 3  
A sample of famous corpora (Pustejovsky& Stubbs, 2013, p.8)

Name of corpus	Year published	Size	Collection contents
British National Corpus (BNC)	1991–1994	100 million words	Cross-section of British English, spoken and written
American National Corpus (ANC)	2003	22 million words	Spoken and written texts
Corpus of Contemporary American English (COCA)	2008	425 million words	Spoken, fiction, popular magazine, and academic texts

**Teaching and learning phonetics:**

As a result, machine learning has made significant contributions to teaching phonetics, most visible in the acoustic feature importance ranking and automatic voice recognition systems, among other applications. Random forests and linear hybrid models were utilized to uncover the most significant prognostic markers in audiology. Al

Tamimi and Khattab (2018) discovered the most predictive markers for differentiating between distinct acoustic stops using random forests and linear mixed models. Arnhold and Kyrolainen (2017) evaluated the concentration scoring by random woodlands and the Universalis added hybrid algorithm, emphasizing the variable significance. They discovered that both algorithms performed admirably in their evaluation. This can support us better to comprehend phonetics and establish a speech grading system that can be utilized in language quizzing, clinical linguistics, and a variety of other settings.



*Figure 6: system speech recognition in ML (Litman et al., 2018, p.295).*

The assistance vector machines, and linear discriminant analysis techniques used by Howell et al. (2017) to train a classifier by listening to various sorts of talks are used in their research. As proven by the examples provided above, machine learning may be applied to phonetics if the acoustic data required to train the model can be converted to a digital format before being used to train the model. In the following step, telephonic researchers can utilize machine learning to construct a model that can be used to solve related problems in the future.



**Answering 2nd question of the research:**

The accuracy or confusion matrix marks can be generated after a classifier or a regressor has been trained and evaluated. We are, on the other hand, occasionally substantially more concerned with the relevance of input factors. There is a problem with the interpretability of machine learning.

Because of the system's inability to be understood by the user, the application of machine learning in applied linguistics, such as autonomous scoring, has been called into question. As a result, the subject of interpretability will be the primary focus of this section. Starting with the random forest algorithm, it's safe to say that it's the most utilized algorithm among ESL/EFL researchers.

One possible explanation for this is that knowledge about critical characteristics can be changed. More importantly, regardless of whether the data follows a normal distribution, it is helpful for most situations. It is still valid for data with many predictors and small samples when the predictors are collinear, as shown in the example (Matsuki et al., 2016). Random forests will also take into consideration the influence of interactions (Baumann & Winter, 2018).

Many language research was conducted using the random forest method. For example, Her and Tang (2020) employed a random forest to rank the importance of characteristics to understand better the prediction capability of input variables in a more accurate manner. With decision trees, the outcome can be seen as a visual representation of the process. This can aid us in better understanding the functionality of the system. For example, by presenting decision trees, Fromont et al. (2020) highlighted the impact of individual variability on the experiment's outcome.

Contrary to decision trees and random forests, neural networks can be more difficult to understand than these models. As a result, specific autonomous grading systems in applied linguistics based on machine learning have been questioned. An alternative strategy, known as Shap value (Ribeiro et al., 2016), can be used to neural networks to describe their behavior. It is possible to define any classifier or regression model with Shap.

On the one hand, linguistic practitioners can validate the automatic scoring system based on deep learning using this way. It can, on the

other hand, provide us with invaluable guidance. Linguists have been slow to adopt this strategy, yet it has been used in other fields of study. The work by Barredo Arrieta et al. provides additional information on explainable ML (2020). Individual conditional expectation and partial reliance plots are also excellent tools for seeing into the black box. The essay by Adadi and Berrada contains additional material (2018). In the future, studies on TESOL/TEFL should make full use of interpretable ML.

### **Answering 3rd question of the research:**

Following a discussion of the apps and interpretability of machine learning, the next problem is how to make it achievable for all linguistic researchers to reach these tools and resources. In answering the third research question, it will be necessary to consider machine learning programming languages, tools, and platforms. The Python programming language should be regarded as the most recommended programming language for machine learning, followed by the R programming language as a starting point for machine learning. After Java and C++, Python is the third most widely used programming language in computer science, owing to its ease of learning and implementation.

It is expected to continue to rise in popularity soon. It is available to the public on a no-fee basis. Also widely used in computer science and artificial intelligence, the R programming language is gaining popularity (Mizumoto & Plonsky, 2016). In the opinion of Hao and Ho (2019), when it comes to the library for implementing machine learning algorithms, Sklearn is without a doubt the most satisfactory candidate because it contains several powerful functions for all types of algorithms.

Keras, based on TensorFlow, is an elastic and robust tool for implementing neural networks in various situations (Pang et al., 2019). Pytorch has also witnessed a surge in the number of users in recent years, particularly among those working in academia. When working with the R programming language, it is advised that you use the RStudio programming environment. Python programming should be used on a platform such as Spyder or Jupyter notebook, supported and recommended by Anaconda.

**Conclusion:**

These are the main contributions made by ML to ESL/EFL learning and teaching, condensed as follows: Aspects of computer-aided language learning and teaching include ID, prognosis, automatic marking progress in (language evaluation), and feature importance rating, amongst other things. Language learning and instruction will become more adaptive and customized due to the use of machine learning techniques. This technology can also assist us in automatically detecting test fraud and providing fine-grained information about a student's ability level in the language. The reliability and validity of language evaluation scores are improved by using automatic scoring based on machine learning.

At long last, machine learning technologies trump classic linear regression methods in terms of accuracy. Even when the assumptions of normality, linearity, and collinearity are violated, complex data types can be handled via techniques such as random forests. Also conceivable is to investigate the interactions that exist among the variables. Future applied linguistics investigations should use interpretable machine learning (ML) to extract more language materials from data.

From the perspective of the query type, ML can provide answers to problems such as regression problems, classification difficulties, clustering difficulties, and dimensionality reduction challenges, among others. It is recommended that researchers in ESL/EFL are learning concentrate more on the types of issues, where the study on the topic of interdisciplinary research between ML and applied linguistics is being conducted jointly.

Most ESL/EFL branches including its teaching and learning appear to be suitable for ML. Almost anything digitalized, such as digital numbers, natural language, images, sound, and other stuff, can be saved in English language learning especially those related to applied linguistics such as phonetics learning. It may also help resolve ESL/ESL learning difficulties as we approach closer to full automation of our life processes. First and foremost, machine learning may continue to assist ESL/EFL researchers in delving into and solving the more challenging linguistic troubles that usual practices cannot solve (Gass et al., 2020). A significant advantage of ML in ESL/EFL learning and teaching over traditional methods is the higher accuracy and flexibility.

**recommendations and suggestions for further research:**

By looking at the conclusions of this study, we can see that: For the most part, there is no apparent boundary between teaching and studying the various sub-fields of the English language, and this is especially true for applied linguistics in higher education. Research attributed to one discipline of applied linguistics can be used in another branch of applied linguistics. Researchers will overcome this limitation if they establish precise and exact requirements for themselves in future experiments. Furthermore, the focus of this study is restricted to a small number of journals devoted to English language teaching and learning.

A substantial number of outstanding works from other fields than applied linguistics may be added to the collection in the future. Some important working research and Ph.D. theses will probably go unnoticed because they are too obscure. Because this research focuses on teaching and learning only a specific subset of the English language, the current study proposes that researchers look at other types of research in addition to the research included here. Furthermore, because the present systematic review does not rule out the possibility of applying machine learning to English language teaching in other fields, future studies may scan all potentially relevant studies to provide a more comprehensive idea of what this review has accomplished.

**References:**

Adadi, A., & Berrada, M. (2018). Peeking inside the Black box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160. <https://doi.org/10.1109/ACCESS.2018.2870052>.

Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J.-D. (2007). Modeling diagnostic assessments with bayesian networks. *Journal of Educational Measurement*, 44(4), 341-359.

Alruily, M. (2021). Classification of Arabic Tweets: A Review, *Electronics*, 10,1-31. <https://doi.org/10.3390/electronics10101143>

Al-Tamimi, J., & Khattab, G. (2018). Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics*, 71, 306-325. <https://doi.org/10.1016/j.wocn.2018.09.010>.

Andoni, A. (2020). Artificial Intelligence Can't Help with the COVID Pandemic. Or can it?

Available at: <https://nodehealth.org/2020/07/08/artificial-intelligence-cant-help-with-the-covid-pandemic-or-can-it/>.

Arnhold, A., & Kyrolainen, A. J. (2017). Modeling the interplay of multiple cues in prosodic focus marking. *Laboratory Phonology*, 8(1), 1–25. <https://doi.org/10.5334/labphon.78>

Ballier, N., Canu, S., Petitjean, C., Gasso, G., Balhana, C., Alexopoulou, T., & Gaillat, T. (2020). Machine learning for learner English: A plea for creating learner data challenges. *International Journal of Learner Corpus Research*, 6(1), 72-103. <https://doi.org/10.1075/ijlcr.18012.bal>.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained german listeners' perceptual judgments. *Journal of Phonetics*, 70, 20-38.

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, 66, 45-60.

Bybee, J., & De Souza, R. N. (2019). Vowel duration in English adjectives in attributive and predicative constructions. *Language and Cognition*, 11(4), 555-581.

Chapelle, C. A., & Chung, Y.-R. (2010). The promise of NLP and speech processing technologies in language assessment. *Language Testing*, 27(3), 301-315.

Charitopoulos, A., Rangoussi, M., & Koulouriotis, D. (2020). On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010-2018. *International Journal of Artificial Intelligence in Education*, 30, 371-430.

Crossley, S. A. (2013). Advancing research in second language writing through computational tools and machine learning techniques: A research agenda. *Language Teaching*, 46(2), 256-271.

Crowther, D., Kim, S., Lee, J., Lim, J., & Loewen, S. (2020). Methodological synthesis of cluster analysis in second language research. *Language Learning*, 71(1), 99-130.

Cui, Y., Gierl, M., & Guo, Q. (2016). Statistical classification for cognitive diagnostic assessment: An artificial neural network approach. *Educational Psychology*, 36(6), 1065-1082.

Deshors, S. C., & Gries, S. T. (2016). Profiling verb complementation constructions across new English: A two-step random forests analysis of *ing* vs. *To* complements. *International Journal of Corpus Linguistics*, 21(2), 192-218.

Fonteyn, L., & Nini, A. (2020). Individuality in syntactic variation: An investigation of the seventeenth-century gerund alternation. *Cognitive Linguistics*, 31(2), 279-308.

Frank, S. L. (2020). Toward computational models of multilingual sentence processing. *Language Learning*, 71(S1), 193-218. <https://doi.org/10.1111/lang.12406>.

Fromont, L. A., Royle, P., & Steinhauer, K. (2020). Growing Random Forests Reveals that Exposure and Proficiency Best Account for Individual Variability in I2 (and I1) Brain Potentials for Syntax and Semantics. *Brain and Language*, 204, 1-17. <https://doi.org/10.1016/j.bandl.2020.104770>.

Gass, S., Loewen, S., & Plonsky, L. (2020). Coming of age: The Past, Present, and Future of Quantitative SLA Research. *Language Teaching*, 1-14. <https://doi.org/10.1017/S0261444819000430>.

Gudmestad, A., House, L., & Geeslin, K. L. (2013). What a Bayesian Analysis Can Do for Sla: New Tools for the Sociolinguistic Study of subject Expression in I2 Spanish. *Language Learning*, 63(3), 371-399.

Hao, J., & Ho, T. K. (2019). Machine learning made easy: A review of the sci-kit-learn package in a python programming language. *Journal of Educational and Behavioral Statistics*, 44(3), 348-361.

Her, O.-S., & Tang, M. (2020). A statistical explanation of the distribution of sortal classifiers in languages of the world via computational classifiers. *Journal of Quantitative Linguistics*, 27(2), 93-113.

Howell, J., Rooth, M., & Wagner, M. (2017). Acoustic classification of focus: On the web and in the lab. *Laboratory Phonology*, 8(1), 16. <https://doi.org/10.5334/labphon.8>.

Hu, Y., & Plonsky, L. (2019). Statistical assumptions in I2 research: A systematic review. *Second Language Research*, 37(1), 171-184. <https://doi.org/10.1177/0267658319877433>.

Hurwitz, J. and Kirsch, D. (2018). *Machine Learning for Dummies*, John Wiley & Sons, Inc.

Kang, H., & Yang, J. (2020). Quantifying perceived political bias of newspapers through a document classification technique. *Journal of Quantitative Linguistics*, 27(1), 1-24. <https://doi.org/10.1080/09296174.2020.1771136>.

Kelly, S., Olney, A. M., Donnelly, P., Nystrand, M., & D'Mello, S. K. (2018). Automatically measuring question authenticity in real-world classrooms. *Educational Researcher*, 47(7), 451-464.

Khany, R., & Tazik, K. (2019). Levels of statistical use in applied linguistics research articles: From 1986 to 2015. *Journal of Quantitative Linguistics*, 26(1), 48-65.

King, K. A., & Mackey, A. (2016). Research methodology in second language studies: Trends, concerns, and new directions. *The Modern Language Journal*, 100(S1), 209-227.

Latifi, S., & Gierl, M. (2020). Automated scoring of junior and senior high essays using coh-Matrix features: Implications for large-scale language testing. *Language Testing*, 38(1), 62–85. <https://doi.org/10.1177/0265532220929918>.

Lei, L., & Liu, D. (2019). Research trends in applied linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics*, 40(3), 540-561.

Lesonen, S., Steinkrauss, R., Suni, M., & Verspoor, M. (2020). Dynamic Usage-based Principles in the Development of L2 Finnish Evaluative Constructions. *Applied Linguistics*, 42(3), 442–472. <https://doi.org/10.1093/applin/amaa030>.

Lin, Z.(2021). A Methodological Review of Machine Learning in Applied Linguistics, *English Language Teaching*, 14(1), 74-85,

Lindstromberg, S. (2016). Inferential statistics in language teaching research: A review and ways forward. *Language Teaching Research*, 20(6), 741-768.

Litman, D., Strik, H., & Lim, G. S. (2018). Speech technologies and the assessment of second language speaking: Approaches, challenges, and opportunities. *Language Assessment Quarterly*, 15(3), 294-309.

Liu, T. (2021). Application of Science and Technology Project Learning in English Linguistics Teaching with Integration of Machine Learning, *J. Phys.: Conf. Ser (The 2nd International Conference on Computing and Data Science)*, 1881(4), 742-750.

<https://doi.org/10.1177/0146621617712246>.

Loukas, S. (2020). What is Machine Learning: Supervised, Unsupervised, Semi-supervised and Reinforcement learning methods, <https://towardsdatascience.com/what-is-machine-learning-a-short-note-on-supervised-unsupervised-semi-supervised-and-aed1573ae9bb>.



Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The random forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading*, 20(1), 20-33.

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092.

<https://doi.org/10.1016/j.jml.2020.104092>.

Mizumoto, A., & Plonsky, L. (2016). R as a lingua franca: Advantages of using r for quantitative research in applied linguistics. *Applied Linguistics*, 37(2), 284-291. <https://doi.org/10.1093/applin/amv025>.

Nicklin, C., & Plonsky, L. (2020). Outliers in l2 research in applied linguistics: A synthesis and data re-analysis. *Annual Review of Applied Linguistics*, 40, 26-55.

Nikitina, L., & Furuoka, F. (2018). Expanding the methodological arsenal of applied linguistics with a robust statistical procedure. *Applied Linguistics*, 39(3), 422-428.

Norouzian, R. (2020). Sample size planning in quantitative l2 research: A pragmatic approach. *Studies in Second Language Acquisition*, 42(4), 849-870.

Norouzian, R., de Miranda, M., & Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning*, 68(4), 1032-1075. <https://doi.org/10.1111/lang.12310>.

Norris, J. M. (2015). Statistical significance testing in second language research: Basic problems and suggestions for reform. *Language Learning*, 65(S1), 97-126. <https://doi.org/10.1111/lang.12114>.

Norris, J. M., Ross, S. J., & Schoonen, R. (2015). Improving second language quantitative research. *Language Learning*, 65(S1), 1-8. <https://doi.org/10.1111/lang.12110>.

Pang, B., Nijkamp, E., & Wu, Y. N. (2019). Deep learning with TensorFlow: A review. *Journal of Educational and Behavioral Statistics*, 45(2), 227-248. <https://doi.org/10.3102/1076998619872761>.

Papi, M., & Teimouri, Y. (2014). Language learner motivational types: A cluster analysis study. *Language Learning*, 64(3), 493-525. <https://doi.org/10.1111/lang.12065>.

Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, 3(1), 61-94. <https://doi.org/10.1075/ijlcr.3.1.03paq>.

Pliakos, K., Joo, S.-H., Park, J. Y., Cornillie, F., Vens, C., & Van den Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education*, 137, 91-103.

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal*, 102(4), 713-731.

Plonsky, L., & Gonulal, T. (2015). Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning*, 65(S1), 9-36. <https://doi.org/10.1111/lang.12111>.

Plonsky, L., & Oswald, F. L. (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition*, 39(3), 579-592.

Pustejovsky, J. & Stubbs, A. (2013). *Natural Language Annotation for Machine Learning*, O'Reilly Media, Inc.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA.

Rico-Juan, J. R., Gallego, A.-J., & Calvo-Zaragoza, J. (2019). Automatic detection of inconsistencies between numerical scores and textual feedback in peer-assessment processes with machine learning. *Computers & Education*, 140, 1-13.

Rudner, L. (2016). Accuracy of Bayes and logistic regression subscale probabilities for educational and certification tests. *Practical Assessment, Research, and Evaluation*, 21(8), 1-11.

Shawar, B. A., & Atwell, E. S. (2005). Using corpora in machine-learning chatbot systems. *International Journal of Corpus Linguistics*, 10(4), 489-516.

Settles, B., LaFlair, G. & Hagiwara, M. (2020). Machine Learning-Driven Language Assessment, *Transactions of the Association for Computational Linguistics*, 8, 247–263. <https://doi.org/10.1162/tacl.a.00310>

Shin, J., & Gierl, M. J. (2020). More efficient processes for creating automated essay scoring frameworks: A demonstration of two algorithms. *Language Testing*, 38(2), 247–272. <https://doi.org/10.1177/0265532220937830>.

Sung, Y. T., Lin, W. C., Dyson, S. B., Chang, K. E., & Chen, Y. C. (2015). Leveling 12 texts through readability: Combining multilevel linguistic features with the CEFR. *Modern Language Journal*, 99(2), 371-391. <https://doi.org/10.1111/modl.12213>.

Th Gries, S. (2020). On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory*, 16(3), 617-647. <https://doi.org/10.1515/cllt-2018-0078>.

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data, *Journal of Phonetics*, 71, 249-267.

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143, 1-18. <https://doi.org/10.1016/j.compedu.2019.103676>.

Wan, M. Y., Fang, A. C., & Huang, C. R. (2019). The discriminativeness of internal syntactic representations in automatic genre classification. *Journal of Quantitative Linguistics*, 28(2), 138-171. <https://doi.org/10.1080/09296174.2019.1663655>.

Wang, R. (2019). Research on Artificial Intelligence Promoting English Learning Change, *Advances in Social Science, Education and Humanities Research*, 325, 392-395.

Warschauer, M., Yim, S., Lee, H., & Zheng, B. B. (2019). Recent contributions of data mining to language learning research. *Annual Review of Applied Linguistics*, 39, 93-112. <https://doi.org/10.1017/S0267190519000023>.

Wiechmann, D., & Kerz, E. (2014). Cue reliance in L2 written production. *Language Learning*, 64(2), 343-364. <https://doi.org/10.1111/lang.12047>.

Wurm, L. H., & Fisiocar, S. A. (2014). What residualized predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, 72, 37-48. <https://doi.org/10.1016/j.jml.2013.12.003>.

Xiao, W., & Sun, S. (2020). Dynamic lexical features of Ph.D. theses across disciplines: A text mining approach. *Journal of Quantitative Linguistics*, 27(2), 114-133. <https://doi.org/10.1080/09296174.2018.1531618>.

Xie, Z. L., Reetzke, R., & Chandrasekaran, B. (2019). Machine learning approaches to analyze speech-evoked neurophysiological responses. *Journal of Speech-Language and Hearing Research*, 62(3), 587-601. [https://doi.org/10.1044/2018\\_JSLHR-S-ASTM-18-0244](https://doi.org/10.1044/2018_JSLHR-S-ASTM-18-0244).

Yang, F., & Li, F. W. B. (2018). Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Computers & Education*, 123, 97-108. <https://doi.org/10.1016/j.compedu.2018.04.006>.

Zhao, X. W., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism*, 13(5), 505-524.

Zheng, Y., Cheon, H., & Katz, C. M. (2020). Using machine learning methods to develop a short tree-based adaptive classification test: A case study with a high-dimensional item pool and imbalanced data. *Applied Psychological Measurement*, 44(7-8), 499-514.