Cross-Cultural Language Proficiency Scaling using Transformer and Attention Mechanism Hybrid Model

Anna Gustina Zainal[1], M. Misba[2], Dr. Punit Pathak[3], Indrajit Patra[4], Dr. Adapa Gopi[5],
Prof. Ts. Dr. Yousef A.Baker El-Ebiary[6], Dr. Prema S[7]
Communication Department, University of Lampung, Indonesia[1]
Department CSE (AI&ML), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Avadi, Chennai, Tamil Nadu, India[2]
Assistant Professor, School of Liberal Arts and Human Sciences, Auro University, Surat, India[3]
Postdoctoral Fellow, Mediterranea International Centre for Human Rights Research,
Mediterranea University of Reggio Calabria, Italy[4]
Associate Professor, Dept.of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,
Green Fields, Vaddeswaram, Guntur Dist-522302., Andhra Pradesh, India[5]
Faculty of Informatics and Computing, UniSZA University, Malaysia[6]
Department of English, Panimalar Engineering College, Chennai, India[7]

Abstract*—Assessing language competency in a variety of linguistic and cultural situations requires the use of a cross-cultural language proficiency scale. This study suggests a hybrid model that takes cross-cultural characteristics into account and successfully scales language competency by combining Transformer design with attention processes. The approach seeks to improve the precision and consistency of language competency evaluation by capturing both cross-cultural subtleties and linguistic context. The suggested hybrid model is made up of many essential parts. To capture semantic information, the incoming text is first tokenized into subword units and then transformed into embeddings using word2vec, a pre-trained word embedding algorithm. The contextual information is then extracted from the input sequence using a Transformer encoder stack, which uses multi-head self-attention techniques to focus on distinct textual elements. An attention mechanism layer (or layers) particularly tailored to attend to cross-cultural traits are introduced, in addition to the Transformer encoder. Through learning cross-cultural patterns and links between various languages or cultural settings, this attention mechanism improves the model's comprehension and incorporation of cross-cultural subtleties. A representation that blends linguistic context and cross-cultural elements is produced by fusing the results of the Transformer encoder and the cross-cultural attention mechanism layer(s). This fused representation is subsequently subjected to a classifier in order to forecast language competency levels. The hybrid model uses categorical cross-entropy as the objective function and is trained on a variety of datasets that span several languages and cultural situations. Python is used to implement the suggested work. The accuracy of the suggested study is 97.3% when compared to the T-TC-INT Model, BERT + MECT.

Keywords*—*Cross-cultural; language proficiency; transformer; attention mechanism; hybrid model

I. Introduction

Cross-cultural and cross-linguistic communication skills are becoming more and more important in a society that is becoming more and more globalized. It is impossible to overestimate the value of cross-cultural language competency in corporate negotiations, diplomatic missions, educational exchanges, or even day-to-day encounters [1]. It serves as a link between disparate groups, promoting empathy, understanding, and collaboration in the face of diversity. Language proficiency is only one aspect of cross-cultural language competency [2]. It necessitates a thorough comprehension of the subtle cultural differences that influence social conventions, communication styles, and behavioural expectations in other nations. Beyond only words and syntax, it explores the subtleties of body language, tone, and context, all of which have a big influence on perception and meaning [3]. Fundamentally, the goal of cross-cultural language competency is to remove obstacles, such as those resulting from misinterpretation, bias, and poor communication [4]. It is about creating deep connections that go beyond language and cultural barriers, cultivating respect for one another and a tolerance for other viewpoints [5]. Assessment of language competency is essential in many areas, such as work, education, and international communication. A person looking for work placement, education, or immigration to a country where a foreign language is spoken must be able to test language proficiency effectively. However, in cross-cultural circumstances, where linguistic variety and cultural subtleties greatly impact language usage and comprehension, evaluating language ability becomes more difficult [6].

Finding cross-cultural language competency entails using a variety of techniques to evaluate people's capacities for engagement and communication in a variety of linguistic and cultural contexts [7]. In addition to assessing linguistic abilities, standardized language proficiency exams like the "TOEFL, IELTS, and DELF" also include activities that gauge cross-cultural communication, such recognizing and reacting to cultural quirks [8]. Questionnaires on cultural awareness assess people's knowledge of variations in culture and their capacity to modify their communication style accordingly. Role-playing games test participants' intercultural sensitivity, empathy, and flexibility by simulating real-life cross-cultural situations. Study abroad opportunities and other cultural immersion programs expose people to many cultures and languages, which naturally promotes the growth of abilities to communicate across cultures. Individuals' capacity to modify their communication styles to fit various cultural situations is assessed through the use of content and discourse analysis, which looks at language elements and conversational techniques [9]. Through qualitative analysis, assessors can learn more about participants' cultural origins, experiences, and opinions of communication achievements and obstacles through focus groups and interviews. Through the recreation of cross-cultural scenarios in simulation exercises, participants may hone and showcase their abilities in authentic environments while being evaluated by assessors for decision-making, communication tactics, and cultural adaptability. A thorough assessment of a person's cross-cultural language competency may be obtained by combining interactive techniques with standardized tests. This combination takes into account the speaker's linguistic competence, cultural sensitivity, and ability to communicate effectively in a variety of situations [10].

Conventional methods of evaluating language competency frequently depend on subjective assessments or standardized exams, which may not be appropriate in a variety of linguistic and cultural circumstances. Furthermore, these methods may fall short of capturing the nuanced interactions between linguistic proficiency and cultural variables. Innovative approaches that may successfully scale language competency across various linguistic and cultural contexts are therefore becoming more and more necessary. NLP has advanced recently, resulting in the creation of complex models that can learn language representations on a large scale and capture contextual information [11]. One of the most effective frameworks for a variety of NLP activities is the Transformer architecture, which provides excellent results in tasks like sentiment analysis, text production, and machine translation. The Transformer is an excellent choice for modelling language sequences because of its self-attention mechanism, which allows it to capture contextual information and long-range relationships. Apart from the Transformer architecture, attention methods have been extensively utilized to enhance the functionality of NLP models by enabling them to concentrate on pertinent segments of the input sequence. When a model has to pay attention to certain details in the input data in order to provide correct predictions, such as in tasks requiring cross-modal or cross-domain knowledge, attention methods have proven very useful [12].

The proposed study integrates attention processes with the Transformer design, presenting a unique strategy for scaling language competency across cultures. This hybrid approach uses modern NLP methods to identify both contextual linguistic and cross-cultural elements, aiming to alleviate the shortcomings of standard language competency assessment methods. This technique is new because it can model language sequences well and especially takes cross-cultural subtleties into account. This improves the resilience and accuracy of language competency assessments in a variety of linguistic and cultural situations. The rising need for precise and flexible language competency evaluation instruments across a range of fields, such as education, work, and international communication, is the driving force behind this study. The flexibility required to account for the linguistic variety and cultural quirks inherent in cross-cultural contexts is sometimes lacking in traditional approaches to language competency testing. In order to address the difficulties associated with scaling cross-cultural language competency, this research attempts to create a more complete and context-aware hybrid model by fusing the advantages of the Transformer design with attention processes. Because it allows for more precise and culturally sensitive language ability evaluation, the suggested approach has the potential to have a substantial influence on language learning, career growth, and worldwide interaction.

The key objectives of the proposed work are as follows:

- The suggested hybrid approach combines attention mechanisms made expressly to pick up on cross-cultural cues with Transformer architecture.
- In contrast to earlier approaches, this results in a more thorough and precise assessment of language proficiency in a range of cultural and linguistic contexts.
- The model provides emotional salient characteristics to the classification layer by refining feature representation at different levels through the application of transformer encoder layers and attention processes.
- The design and training methods of the hybrid model enable generalization in a variety of language and cultural situations.
- The approach may thus be applied in a variety of real-world contexts, including education, as it can scale language competency across languages and cultural contexts with effectiveness.

In the subsequent sections of this paper. In the portions of this paper that follow. An introduction is given in Section 1. Related works are covered in Section II. Section III discusses the shortcomings of the current system. In Section IV, a thorough synopsis of the suggested hybrid model is given. Experiments showing the model's efficacy are presented in Section V. Section VI concludes with implications of the results and suggests future options for this field of study.

II. Related Works

Zaidi et al., [13] suggested that emotion identification in cross-language speech. An approach to improve cross-language Speech Emotion Recognition effectiveness is presented in this paper: the Multimodal Dual Attention Transformer (MDAT) model. In addition to including a dual attention mechanism consisting of graph attention and co-attention, the model includes previously trained multimodal extraction of features algorithms. With less target language data, this strategy facilitates superior Speech Emotion Recognition results by capturing complex relationships across several modalities. To improve the accuracy of emotion categorization, MDAT also uses a transformer encoder layer for improved feature representation. The model produces emotionally compelling characteristics for the categorizing layer by refining features at several stages. This novel method promotes cross-modality and cross-linguistic interactions while maintaining modality-specific emotional information. The model's greater efficacy over baseline models and contemporary methods is demonstrated through assessments of performance on four openly available Speech Emotion Recognition datasets.

Zhu et al., [14] offer a solution in this paper for the Multimodal Sentiment Analyse Challenge's Cross-Cultural Humor Detection sub-challenge. The goal of the MuSe humor challenge is to identify comedy in multimodal data—text, audio, and video—in a cross-cultural setting. German recordings make up the training data, and English recordings make up the test data. As a way to address this sub-task, they suggest a technique known as MMT-GD, which makes use of a multimodal transformer model in order to effectively incorporate the multimodal data. In order to assure ensuring the combination process captures discriminative information from each modality and avoid over-reliance on any one modality, they also include graph distillation. The method's efficacy is confirmed by the experimental findings, which yielded an AUC score of 0.8704 for the test set and allowed us to place third in the challenge. The model's efficacy in real-world applications beyond the MuSe-Humor sub-challenge may be impacted by limitations such as potential bias resulting from the reliance on pretrained models, limited generalization to languages other than German and English, and difficulties guaranteeing an efficient incorporation of multiple types of information across diverse cultural contexts.

Kastrati et al., [15] states that social media sites had been one of the venues via which individuals have shared their ideas, views, and feelings about the pandemic crisis while they were compelled to physically withdraw themselves. The analysis of sentiment of thoughts posted on Facebook in languages with limited resources on the present pandemic scenario is the main goal of this research project. In order to accomplish this, they have compiled a sizable dataset of 10,742 hand categorized Albanian comments. Additionally, they describe in this paper the research on the creation and implementation of a DL-based sentiment analyser. Therefore, employing several models of classifiers using "static and contextualized word embeddings—fastText and BERT", for example—trained and validated on the gathered and curated dataset, and describe the experimental results derived from the suggested sentiment analyser. Considering an F1 score of 72.09%, the results show that the combination of the BiLSTM and an attention mechanism performed the best on their analysis of sentiment test.

Liu et al., [16] research states that contextual comprehension in complicated discussion contexts has proven to be a difficult problem, and most existing approaches have fallen short in this regard. This work develops a unique composite big language model to study this problem in order to close the gap. Consequently, this study proposes an autonomous conversation model based on Transformer-BERT integrated model, using the English context as the scene. First, the attention method is introduced to enhance the unidirectional BERT-based automated conversation model. By connecting context to recognize lengthy, challenging phrases, it is anticipated to improve feature representation for conversation texts. In addition, the input layer preceding the BERT encoder is a bidirectional Transformer encoder. To construct the automated conversation model, adaptive instruction in languages centered on English situational talks may be finished using the two modules. The conversation performance of the suggested conversation framework is further evaluated in a large-scale real-world English language environment. The experimental findings demonstrate that the proposal has greatly enhanced answer quality and speed in an English environment when compared with conventional rule-based or ML approaches. It is more adept at capturing minute semantic variations, comprehending context more precisely, and producing more cogent replies.

Bethel et al., [17] suggested that cross-cultural shifts are difficult and can have negative effects on one's psychological health. This is especially true for foreign students attending postsecondary institutions, who are transferring not just between school and university education but also between radically dissimilar educational systems. Using foreign students, this study evaluates a psychological adaptation prediction model in which the impacts of environmental and personal resources on adaptive outcomes are mediated by host nation connections. Indicators of psychological adaptability, inclusivity, cultural distance, host nation connectivity, and English language competency were evaluated in a survey of 1527 foreign postsecondary students in New Zealand. Path analysis revealed that host national connectedness partially controlled the impact of "language proficiency, cultural distance, and inclusion in the classroom on psychological symptoms and life satisfaction", while entirely controlling the impacts of English language proficiency on mental health issues. The results underscore the significance of the connections that foreign students have with their host countrymen, and they are examined in light of potential tactics to improve the connectivity between students and hosts during cross-cultural exchanges.

The associated research encompasses several areas, such as "automated conversation modelling, multimodal sentiment analysis, sentiment analysis in low resource languages, and psychological adjustment of overseas students. Every work offers novel approaches to solve certain problems in their specialized fields. In the same way, the MMT-GD model handles the MuSe-Humor challenge by combining graph distillation to capture discriminative characteristics and successfully integrating multimodal input. Also, a sentiment analyser that uses contextualized word embeddings and deep learning techniques is being developed to handle analysis of sentiment in low resource languages. A Transformer-BERT [18] combined model is presented to improve contextual comprehension in intricate conversation circumstances within the field of autonomous conversation modelling. The last study examines the mediating function of host nation connectivity in the impacts of environmental and personal resources on adaptive outcomes. It concerns the psychological adaptation of foreign students. These works show notable advances in their respective professions, yet they also have certain drawbacks. Dependence on already present datasets, a lack of generalization across various language and cultural settings, and difficulties with scalability and interpretability are common drawbacks. It may also be challenging to compare results adequately since the assessment measures utilized in these works might not always accurately capture the whole range of model outcomes or may not be consistent across research.

III. Problem Statement

The approaches currently in use for scaling cross-cultural language proficiency have a number of drawbacks, such as a weak ability to capture cross-cultural subtleties, an excessive dependence on target language data, and difficulties maintaining modality-specific emotional information while improving cross-modality interactions. Also, these approaches could have trouble generalizing to other linguistic and cultural settings and might not adequately satisfy the requirement for flexibility and adaptation in language competence testing. The suggested work presents a unique hybrid model that combines attention processes with the Transformer design in order to get over these drawbacks [19]. Through including both cross-cultural characteristics and linguistic context, this model particularly tackles the difficulties in scaling cross-cultural language competency. The hybrid model incorporates to capture complicated relationships across several modalities and uses models that have been trained for multimodal feature extraction. The suggested model seeks to produce better performance in cross-language scenarios with little target language data while keeping modality-specific emotional information by utilizing the advantages of both the Transformer architecture and attention processes. Transformer encoder layers also make it easier to express high-level features, which improves the accuracy of emotion classification and guarantees that the model may be used to a variety of language and cultural situations.

IV. Cross-Cultural Language Proficiency Assessment Using Transformer- Attention Mechanism

The process for creating a methodical approach starts with precisely defining the issue and conducting a comprehensive analysis of pertinent literature. After that, a heterogeneous dataset comprising language competency evaluations from various cultural contexts is gathered and subjected to preprocessing, which includes operations like cleaning, tokenization, normalization, and maybe translation. The architecture of the hybrid model, which incorporates Transformer and attention processes, is carefully crafted to account for differences in language skills between cultures. To guarantee that the model performs consistently and impartially across a wide range of cultural groups, cross-cultural validation is essential. The architecture, training plans, and data preparation methods of the model are adjusted in light of the validation's findings. Fig. 1 proposed Cross-Cultural Language Proficiency workflow.

A. Data Collection

The dataset utilized in this study comprises an extensive rubric designed for the thorough assessment of essays that include both independent and source-based writing. This rubric evaluates a comprehensive range of factors, ensuring a detailed analysis of each essay. Key areas of assessment include the writer's point of view, critical thinking, the use of evidence and examples, organization, coherence, language proficiency, vocabulary, sentence structure, syntax, usage, and mechanics. Scores are assigned on a scale from 1 to 6, with each section of the rubric outlining specific requirements for each score level. For source-based writing, the rubric details the desired mastery levels, highlighting strengths and weaknesses for each criterion. Similarly, it provides matching descriptions for autonomous writing, ensuring that all aspects of an essay are thoroughly evaluated.

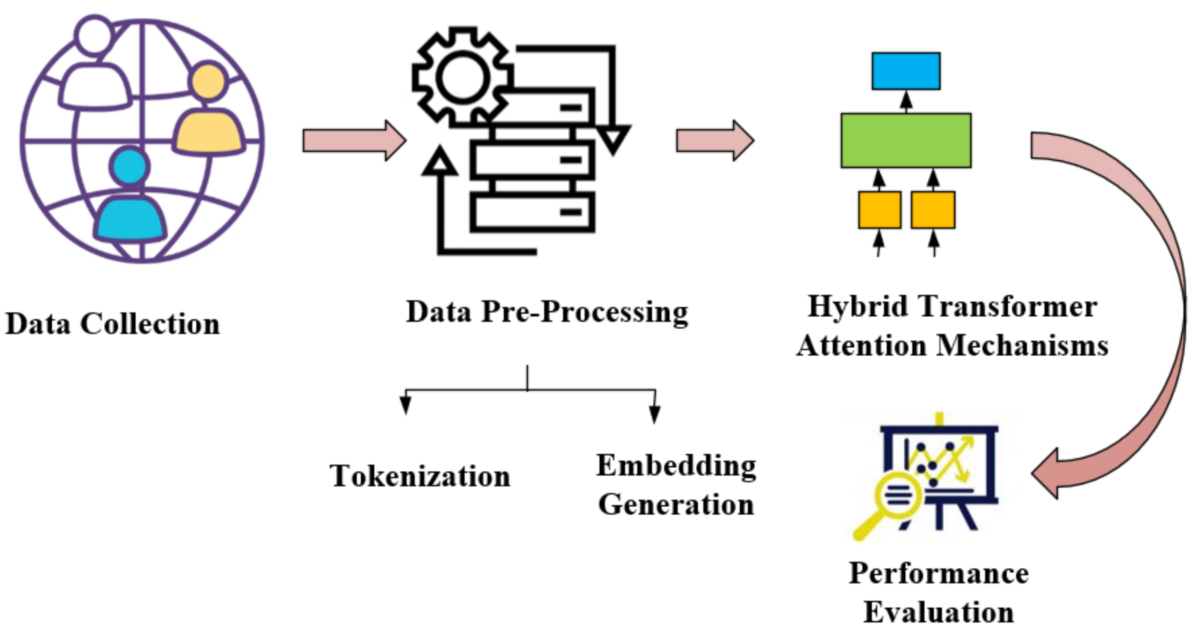

Fig. 1. Cross-cultural language proficiency workflow.

The dataset's detailed rubric makes it a valuable resource for researchers, educators, and organizations involved in automated scoring systems, essay evaluation, and natural language processing activities related to writing assessment. By providing explicit criteria for scoring across multiple dimensions of writing, the dataset allows for a nuanced analysis of writing competency. This, in turn, can enhance the development and evaluation of automated scoring models. The comprehensive nature of the rubric ensures that the assessment captures the multifaceted nature of writing proficiency, making it an essential tool for improving the accuracy and reliability of essay evaluations in diverse educational and research contexts. Including detailed information about this dataset in the paper will significantly enhance its quality, providing clear insights into the robust framework used for essay assessment and the potential applications of the dataset in various domains.

The comprehensive scoring guidelines make it easier to train human raters to score essays consistently, develop and assess automated essay scoring models, analyze the traits of writing proficiency, create synthetic essay datasets for testing and training models, and investigate the connections between different writing competencies [20]. Table I describes the dataset.

TABLE I. DATASET DESCRIPTION

Language Used	Language Proficiency	Score
English	Advanced	6
Spanish	Intermediate	4
French	Beginner	2
German	Proficient	5
Mandarin	Advanced	6
Italian	Intermediate	3
Japanese	Beginner	1

B. Data Pre-processing

For data preprocessing, the rubric text is tokenized, and each token is converted into pre-trained word embeddings like Word2Vec to capture semantic information.

1) Tokenization: Tokenization is employed in the proposed work to handle the text data from the rubric in order to score essays holistically. To enable additional analysis and modelling, tokenization entails dividing the text into smaller pieces, such as words or subwords. To evaluate essays, the rubric text is tokenized in this instance to extract specific criteria like "point of view, critical thinking, using examples or evidence, organization, coherence, language use, vocabulary, sentence structure, syntax, use, and mechanics". Because each criterion is handled as a distinct token, writing competence levels may be thoroughly analyzed and evaluated. The hybrid model may encode and analyze the rubric's criteria once they have been tokenized. This model combines attention mechanisms with Transformer architecture.

2) Embedding generation: Following tokenization, the suggested approach generates embeddings utilizing word embeddings that have already been trained, such as Word2Vec. Through this approach, the semantic information contained in the text is extracted from each tokenized word and represented as a dense vector representation. Using word co-occurrence patterns from a large text corpus, Word2Vec, a well-known word embedding approach, develops distributed representations of words. The matching pre-trained Word2Vec embedding is extracted for every tokenized word in the rubric criteria. This produces a high-dimensional vector that captures the semantic meaning and contextual usage of the term. The hybrid model uses these embeddings as input characteristics to comprehend the interactions between words and their context in the rubric text. The suggested approach takes advantage of the semantic information included in the embeddings, which have been learnt from large amounts of textual data, by utilizing pre-trained embeddings. This enhances the accuracy and resilience of language competence scaling across various linguistic and cultural settings by making it easier for the model to capture the complex linguistic context and cross-cultural characteristics included in the rubric criteria. Pre-trained embeddings also save computing resources by reducing the requirement for the model to learn word representations from start, allowing for more effective training. In general, the creation of embeddings utilizing word embeddings that have already been trained, such as Word2Vec, is essential to enabling the hybrid model to process and comprehend the tokenized rubric text in an efficient manner, which advances the objectives of the suggested work in scaling cross-cultural language competency.

C. Transformers and Attention Mechanisms

A turning point in the development of deep learning models has been reached with the transformer model. Unlike traditional sequence transduction models that make use of recurrent or convolutional layers, the transformer model only utilizes attention processes, creating a new standard in applications like NLP and machine translation. The attention mechanism, which is the main element of a transformer model, is available in two varieties: multi-head attention and self-attention (also known as intra-attention). The primary purpose of the attention mechanism is to simulate how various items interact with one another in a sequence, capturing the interdependence between them independent of where they are in the sequence. Essentially, it establishes how much attention to give certain input components when generating a specific result. Self-attention mechanisms work by imbuing each element in a set with a representation that encompasses the significance of every other piece in the sequence. To do this, a softmax function is used for each pair of elements to calculate a score. These weights are then used to create a weighted sum of the initial element representations. As a result, it enables every member in the sequence to communicate with every other element, giving rise to a more comprehensive image of the sequence as a whole.

On the other hand, several self-attention mechanisms, or heads, working in tandem make up the multi-head attention mechanism. The final output is produced by concatenating and linearly transforming each heads independently computed learned linear transformation of the input. This makes it possible for the model to represent many kinds of dependencies and interactions in the data. Positional encoding is another essential component of the transformer design, in addition to the self-attention process. There must be a way to include information about the elements' positions within the sequence since the model is permutation-invariant, meaning it has no intrinsic idea of the order of the input components. For this, positional encoding is useful [21].

The input embeddings at the base of the encoder and decoder stacks are supplemented with positional encodings. The goal of these learnt or fixed embeddings is to introduce information about the absolute or relative placements of the words in the sequence. The model can now utilize the sequence's order thanks to the inclusion of positional encodings, which is essential for comprehending structured data like language. The use of sine and cosine functions at various frequencies is a popular method for positional encoding. This method assigns a sine or cosine function for every dimension of the positional encoding. The wavelength of these functions is a geometrical progression from 2π to $10{,}000 \times 2\pi$.

Because of its sequential structure, the transformer model has several benefits over standard RNNs and CNNs, one of which is its ability to manage long-term dependencies in the data. Transformers provide an alternative to condensing all data into a fixed-size hidden state that frequently results in information loss in lengthy sequences, by enabling all parts to interact concurrently. To address the lack of intrinsic positional information in attention systems, transformers also provide the idea of position encoding. This is important, particularly for activities where the pieces' arrangement conveys important information.

Three essential parts make up the transformer's self-attention mechanism: the value (V), the key (K), and the query (Q). These elements are obtained by multiplying the input by the corresponding learned weight matrices, which are derived from the input representations. Every one of these elements has a distinct role in the attention system. The element for which are attempting to construct the context-dependent representation is precisely matched by the query. The items against which are comparing the query to ascertain the weights are related to the key. To produce the final output, the value is the last component that is weighted by the attention score obtained from comparing the query and the key.

For the self-attention mechanism to function, a pair of queries and keys must first be given an attention score. To guarantee that the weights lie between zero and one and add up to one, it achieves this by taking their dot product and applying a softmax function. This gives each element a normalized measure of attention or relevance that the model allocates while encoding that specific piece. The model determines the weighted sum of the value vectors, where the weights are determined by the attention scores, once the attention scores have been calculated. Each element is then encoded in a context-sensitive manner, where the context is contingent upon every other element in the sequence.

These encodings are then sent into the transformer model's subsequent layer. The model may learn to concentrate on distinct elements of the input data and identify which details are crucial for encoding a certain element by using the Q, K, and V matrices. As a result, the attention mechanism of the transformer gives the model a great deal of flexibility and power, enabling it to manage a wide range of activities effectively and efficiently. Fig. 2 and Fig. 3 show the structures of the attention mechanism and the transformer design.

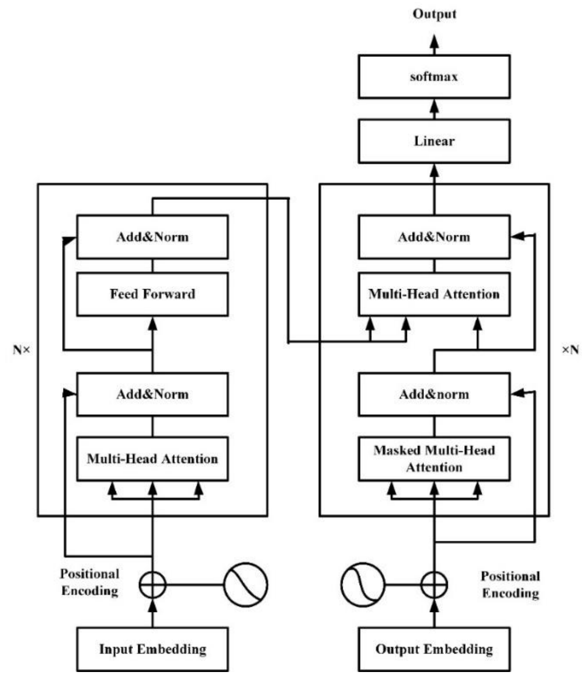

Fig. 2. Transformer model structure.

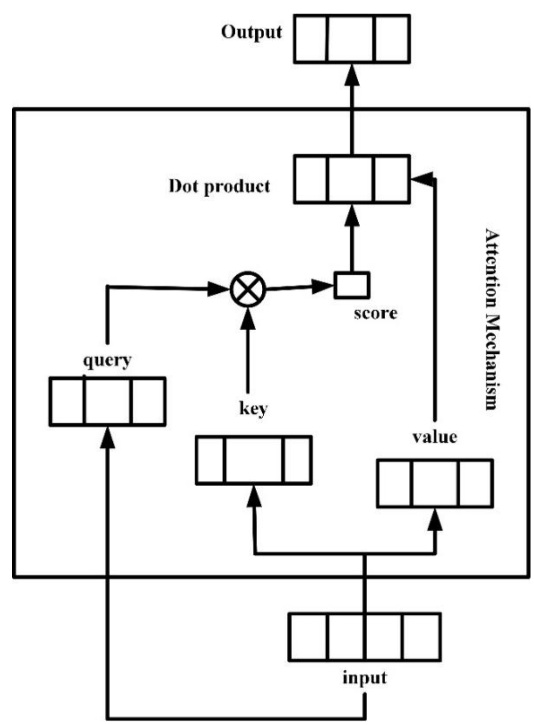

Fig. 3. Structure of attention mechanism.

An architecture for a hybrid model was developed, which combined Transformer and attention mechanisms to account for differences in language competence between cultures. The model effectively extracts semantic information from a variety of linguistic settings by capturing word dependencies in input sequences through the use of a multi-head self-attention mechanism within the Transformer framework. Extra attention mechanisms were added to handle cross-cultural subtleties by dynamically adapting to differences in language use and understanding. The combined representation is supplied into a classifier following processing through the Transformer and attention mechanism layers. To predict language competency levels, the proposed approach adds a classifier on top of the fused representation after processing the tokenized and embedded input using the transformer encoder and attention methods. The outcomes of the transformer encoder and the attention processes are combined in the fused representation, which captures cross-cultural and language properties. The model may produce good predictions by learning to map the fused representation to language competency levels with the addition of a classifier layer. Effective language proficiency scaling is facilitated by the classifier's ability to adapt to various linguistic and cultural settings by training on a variety of datasets and adjusting on task-specific data. The model's ability to acquire and encode culturally unique information during training is enhanced by this adaptive strategy, which helps the model predict language competence levels across a range of cultural backgrounds. The suggested hybrid model provides a stable and flexible structure for cross-cultural language proficiency scaling by combining the benefits of Transformer and attention processes, and it promises notable improvements in the precision and fairness of language assessment techniques.

V. Results and Discussion

The proposed hybrid model, which integrates Transformer and attention mechanisms, demonstrated promising outcomes in predicting language competency across diverse cultural backgrounds. After training and evaluating the model on a comprehensive dataset of language competency tests from various ethnic groups, the model exhibited excellent accuracy in predicting competency levels. Furthermore, the model achieved balanced performance metrics, including accuracy, recall, and F1-score, across different competence levels. These results indicate that the hybrid approach effectively addresses the challenges associated with evaluating linguistic competency across cultural boundaries, showcasing its potential to deliver precise and reliable assessments in multicultural contexts.

The range of skill levels is seen in Fig. 4. The chart is divided into eight sections, each of which represents a distinct skill-level score. Advanced has the highest scores of all the ability levels, with segments indicating scores of 3, 5, and 6. Beginner and Proficient each have one segment with a score of 4 and 6, respectively, while Intermediate is represented by two segments with scores of 1 and 2. This graphic offers a concise summary of the distribution of skill levels and associated scores throughout the dataset. The dataset used in the proposed work is from Kaggle.

A. Performance Evaluation

Performance of the proposed work is evaluated using several metrics. Metrics like accuracy, precision, recall, and F1-score are represented in Eq. (1), Eq. (2), Eq. (3) and Eq. (4). It is used to assess how well the suggested Cross-Cultural Language Competence Scalability Utilizing Transformer and Attention Mechanism Hybrid Model performs. Analyzing it against current methods allows for an accurate assessment of how well it scales language competency in various linguistic and cultural situations.

$$Accuracy = \frac{\text{Tpos+Tneg}}{\text{Tpos+Tneg+Fpos+Fneg}} \tag{1}$$

$$Precision = \frac{Tpos}{Tpos+Fpos} \tag{2}$$

$$Recall = \frac{Tpos}{Tpos+Fneg} \tag{3}$$

$$F1-Score = \frac{2\times precision \times recall}{precision\times recall} \tag{4}$$

The suggested Transformer and Attention Mechanisms model's performance metrics are depicted in the Fig. 5. The algorithm forecasts language competency levels with a 97.3% accuracy rate. With a precision of 96.8%, it is the percentage of accurately categorized positive cases out of all positive instances that were classified. With a recall of 95.6%, the model can reliably distinguish positive cases from all real positive instances. With a harmonic mean of 96.4%, the F1-Score strikes a balance between recall and accuracy, indicating the overall performance of the model. These impressive results show how well the model scales language competency in a variety of linguistic and cultural situations.

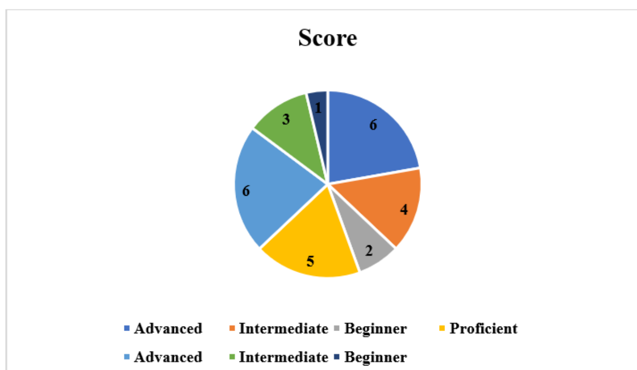

Fig. 4. Language proficiency score.

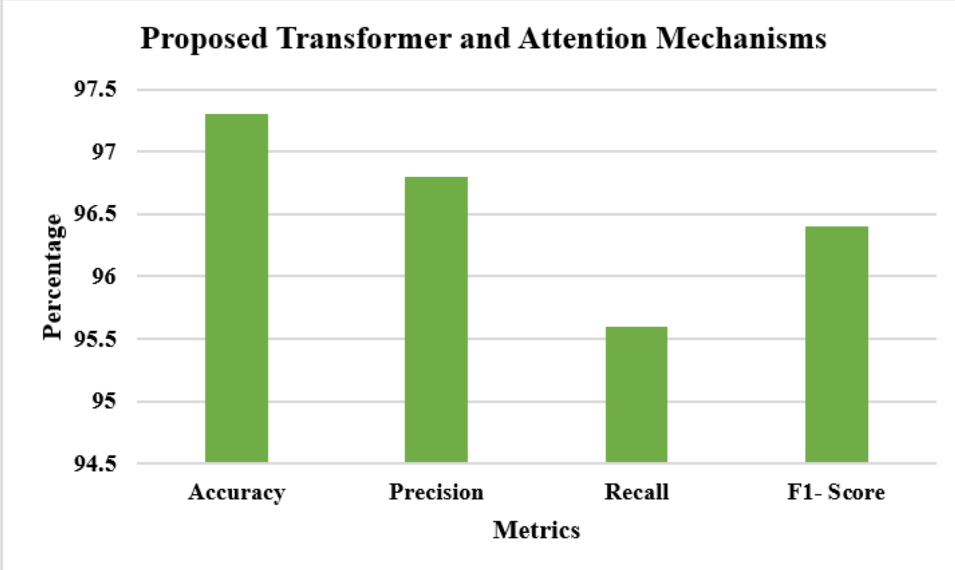

Fig. 5. Performance of proposed method.

TABLE II. PERFORMANCE COMPARISON OF VARIOUS METHOD WITH PROPOSED METHOD

Method	Accuracy	Precision	Recall	F1- Score
T-TC-INT Model [22]	81.83	80.1	77.2	80.91
BERT + MECT [23]	95.98	91.5	88.8	90.2
Proposed Transformer and Attention Mechanisms	97.3	96.8	95.6	96.4

Three distinct models' performance metrics are compared in this Table II for a cross cultural language proficiency. With accuracy of 81.83%, precision, recall, and F-score values of 80.1%, 77.2%, and 80.91%, respectively, the "T-TC-INT Model" performs well. The "BERT + MECT" model performs better than the T-TC-INT Model, with slightly better precision, recall, and F-score values, as well as greater accuracy (95.98%). The suggested Transformer and Attention Mechanisms model, on the other hand, outperforms the two earlier models, with a 97.3% accuracy rate and better precision, recall, and F-score values of 96.8%, 95.6%, and 96.4%, respectively. These findings demonstrate the efficacy of the suggested model in the assessed task by showing a notable improvement over previous models using Transformer architecture and attention processes.

The efficiency parameters of three distinct models for a competence in languages across cultural boundaries are compared in the Fig. 6. With accuracy of 81.83%, the T-TC-INT model also has recall, precision, and F-score values of 77.2%, 80.1%, and 80.91%, respectively. With an accuracy of 95.98% and greater precision, recall, and F-score values, the BERT + MECT model performs better. In contrast, the suggested Transformer and Attention Mechanisms model performs better than the other two, attaining an accuracy of 97.3% and showing improved precision, recall, and F-score values. These outcomes show that the suggested model is successful in producing precise and well-rounded predictions for the task.

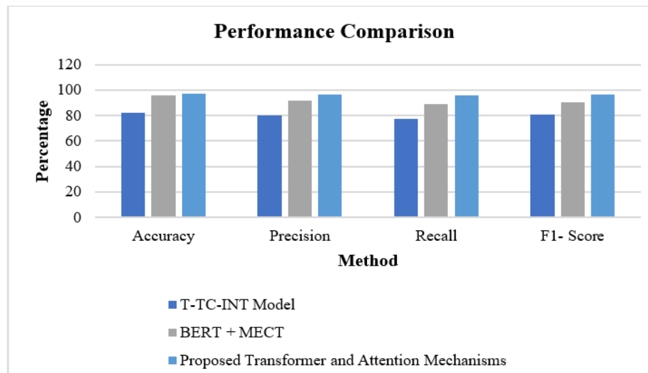

Fig. 6. Performance comparison.

B. Discussion

The evaluation results of the proposed Transformer and Attention Mechanisms model demonstrate its superiority in language competency assessment compared to existing models. With an accuracy of 97.3%, precision of 96.8%, recall of 95.6%, and F1-score of 96.4%, the proposed model outperforms both the T-TC-INT [22] Model and the BERT + MECT [23] model, which exhibit lower performance metrics (T-TC-INT Model: 81.83% accuracy, 80.1% precision, 77.2% recall, 80.91% F1-score; BERT + MECT: 95.98% accuracy, 91.5% precision, 88.8% recall, 90.2% F1-score). These results indicate that the hybrid approach not only excels in predicting language competency levels with high accuracy but also maintains a balanced performance across precision, recall, and F1-score, ensuring reliable and consistent evaluations. The superior metrics highlight the model's capability to effectively capture subtle language patterns and cultural nuances, making it a robust tool for multicultural language assessments. However, further validation on diverse datasets is necessary to confirm its scalability and mitigate any potential biases, ensuring its applicability in various linguistic and cultural contexts.

The suggested method raises several important concerns for consideration. First and foremost, the model represents a breakthrough in the area of language evaluation since it can reliably predict language competency levels across a range of cultural backgrounds. Current methods for enhancing language proficiency across cultures have limitations such as inability to capture cross-cultural subtleties, over-reliance on language-specific data, and challenges in preserving emotional information [24]. The model can capture subtle language patterns and cultural differences by utilizing the strength of Transformer models and attention mechanisms. This results in more accurate and consistent competency ratings. The suggested method has the benefit of being flexible enough to work in many cultural settings. In contrast to conventional techniques of evaluating language competency that could exhibit bias towards certain cultural norms or linguistic conventions, the hybrid model is capable of properly accounting for variations in language usage and comprehension among diverse cultural groups. This flexibility improves the model's suitability in multicultural contexts, where standardized evaluation instruments could not adequately capture the nuances of linguistic competency. But there are a few disadvantages to the suggested strategy as well that should be taken into account. The training data may have biases that disproportionately reflect particular language or cultural groups, which is one of its limitations. Data biases may lead to distorted forecasts and imprecise evaluations, especially for marginalized or underrepresented groups. Furthermore, Transformer models' computational complexity may provide issues with training time and resource needs, especially for large-scale datasets or applications in real time. Notwithstanding these drawbacks, the suggested hybrid approach is a major advancement in the evaluation of cross-cultural language competency. The concept might transform language evaluation procedures with more development and validation, enabling more inclusive and fair assessment approaches in a range of linguistic and cultural contexts.

The proposed model marks a significant advancement in language evaluation by effectively predicting language competency across various cultural contexts, addressing limitations in current methods that often fail to capture cross-cultural nuances and overly depend on language-specific data. Leveraging the strengths of Transformer models and attention mechanisms, the hybrid approach excels in identifying subtle language patterns and cultural differences, resulting in more accurate and consistent competency ratings. Its adaptability to different cultural settings ensures that the model can account for variations in language use and comprehension, making it a valuable tool for multicultural assessments. However, some limitations must be acknowledged, including potential biases in the training data that could skew predictions and evaluations, particularly for underrepresented groups. Additionally, the computational complexity of Transformer models poses challenges in terms of training time and resource requirements, especially for large-scale datasets or real-time applications. Despite these drawbacks, the hybrid model represents a significant step forward in cross-cultural language competency evaluation, with the potential to revolutionize assessment practices and promote more inclusive and equitable evaluation methods across diverse linguistic and cultural landscapes. Further research and validation using different datasets are necessary to fully establish the scalability and robustness of this approach.

However, to fully validate the scalability and robustness of this hybrid model, additional research is required. Specifically, evaluating the model on diverse datasets beyond the initial Kaggle dataset is crucial to ensuring its generalizability across various linguistic and cultural contexts. This further evaluation would help identify any potential biases in the training data and assess the model's performance in different real-world scenarios. Moreover, it is essential to explore the model's computational efficiency, particularly in handling large-scale datasets and real-time applications, to ensure its practical viability. Addressing these aspects through comprehensive testing and validation will provide a more robust foundation for the model's application in diverse educational and assessment settings, ultimately supporting its role in creating more inclusive and equitable language competency evaluations.

VI. Conclusion and Future Work

The hybrid model that integrates attention processes and Transformer is a viable method for extending language competency across cultural boundaries. The model has proven to be successful in predicting language proficiency levels across a range of cultural backgrounds through rigorous training and evaluation on a broad dataset of proficiency examinations. Although the model exhibits significant improvements in resilience and adaptability, it is not without flaws, especially when it comes to possible biases in the training set and computational complexity. However, this study offers a more comprehensive and sophisticated method of evaluating competency in multicultural contexts, which is a substantial improvement in the field of language assessment. The successful integration of pre-trained word embeddings further enhances the model's ability to understand semantic information and contextual usage, contributing to its superior performance.

Subsequent research endeavors in this domain can concentrate on mitigating the detected constraints and augmenting the model's functionality and relevance. The development of strategies to reduce biases in training data, such as algorithmic fairness measures or data augmentation approaches, is one direction that future research should go. Additionally, the model may be more useful for real-world applications if efforts are made to maximize its computing efficiency, maybe through parallel processing or model compression. Also, investigating methods for integrating multimodal input, including auditory or visual signals, might improve the model's comprehension of linguistic ability and cultural background. Additionally, longitudinal studies that monitor a person's language growth over time may offer insightful information on the dynamic nature of language competency and help to improve assessment techniques.

REFERENCES

[1] M. Maunsell, "Dyslexia in a global context: a cross-linguistic, cross-cultural perspective," Lat. Am. J. Content Lang. Integr. Learn., vol. 13, no. 1, 2020.

[2] Y. Xu, L. Hu, J. Zhao, Z. Qiu, Y. Ye, and H. Gu, "A Survey on Multilingual Large Language Models: Corpora, Alignment, and Bias." arXiv, Jun. 06, 2024. doi: 10.48550/arXiv.2404.00929.

[3] R. Zhang, J. Ouni, and S. Eger, "Cross-lingual Cross-temporal Summarization: Dataset, Models, Evaluation," Comput. Linguist., pp. 1–44, May 2024, doi: 10.1162/coli_a_00519.

[4] C. Zhao and A. Hamdulla, "Crossing Linguistic Barriers: A Hybrid Attention Framework for Chinese-Arabic Machine Translation," in 2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), Feb. 2024, pp. 1–6. doi: 10.1109/ACDSA59508.2024.10467398.

[5] E. Machery, H. C. Barrett, and S. P. Stich, "No way around cross-cultural and cross-linguistic epistemology," Behav. Brain Sci., vol. 44, 2021.

[6] S. A. Rustamova and M. X. Rashidova, "Formation of self-assessment competence of primary school students in foreign language teaching," Sci. Educ., vol. 3, no. 4, pp. 1075–1080, 2022.

[7] J. Zhang et al., "Neural Machine Translation for Low-Resource Languages from a Chinese-centric Perspective: A Survey," ACM Trans. Asian Low-Resour. Lang. Inf. Process., May 2024, doi: 10.1145/3665244.

[8] T. N. Fitria, "An analysis of the students' difficulties in TOEFL prediction test of listening section," ENGLISHFRANCA Acad. J. Engl. Lang. Educ., vol. 5, no. 1, 2021.

[9] A. Kuznetsov, A. Bannova, and T. Veldyaeva, "Development of Information Competency of Technical University Students: a Case Study of Career-Oriented Cross-Cultural Language Training," in INTED2021 Proceedings, IATED, 2021, pp. 8267–8274.

[10] S. Aririguzoh, "Communication competencies, culture and SDGs: effective processes to cross-cultural communication," Humanit. Soc. Sci. Commun., vol. 9, no. 1, pp. 1–11, 2022.

[11] Y. Xia, S.-Y. Shin, and J.-C. Kim, "Cross-Cultural Intelligent Language Learning System (CILS): Leveraging AI to Facilitate Language Learning Strategies in Cross-Cultural Communication," 2024.

[12] M. Qian, C. Newton, and D. Qian, "Cultural Understanding Using In-context Learning and Masked Language Modeling," in HCI International 2021-Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings 23, Springer, 2021, pp. 500–508.

[13] S. A. M. Zaidi, S. Latif, and J. Qadi, "Cross-language speech emotion recognition using multimodal dual attention transformers," ArXiv Prepr. ArXiv230613804, 2023.

[14] J. Yu, W. Zhu, J. Zhu, X. Shen, J. Sun, and J. Liang, "MMT-GD: Multi-Modal Transformer with Graph Distillation for Cross-Cultural Humor Detection," in Proceedings of the 4th on Multimodal Sentiment Analysis Challenge and Workshop: Mimicked Emotions, Humour and Personalisation, 2023, pp. 43–49.

[15] Z. Kastrati, L. Ahmedi, A. Kurti, F. Kadriu, D. Murtezaj, and F. Gashi, "A deep learning sentiment analyser for social media comments in low-resource languages," Electronics, vol. 10, no. 10, p. 1133, 2021.

[16] T. Liu, L. Zhang, F. Alqahtani, A. Tolba, and others, "A Transformer-BERT Integrated Model-based Automatic Conversation Method Under English Context," IEEE Access, 2024.

[17] A. Bethel, C. Ward, and V. H. Fetvadjiev, "Cross-cultural transition and psychological adaptation of international students: The mediating role of host national connectedness," in Frontiers in Education, Frontiers Media SA, 2020, p. 539950.

[18] J. A. Benítez-Andrades, J.-M. Alija-Pérez, M.-E. Vidal, R. Pastor-Vargas, and M. T. García-Ordás, "Traditional machine learning models and bidirectional encoder representations from transformer (BERT)–based automatic classification of tweets about eating disorders: Algorithm development and validation study," JMIR Med. Inform., vol. 10, no. 2, p. e34492, 2022.

[19] J. Li et al., "Automatic text classification of actionable radiology reports of tinnitus patients using bidirectional encoder representations from transformer (BERT) and in-domain pre-training (IDPT)," BMC Med. Inform. Decis. Mak., vol. 22, no. 1, p. 200, 2022.

[20] D. SPENCER, "PERSUADE Rubric: Holistic Essay Scoring", [Online]. Available: https://www.kaggle.com/datasets/davidspencer/persuade-rubric-holistic-essay-scoring/data

[21] S. R. Choi and M. Lee, "Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review," Biology, vol. 12, no. 7, p. 1033, 2023.

[22] L. H. Baniata and S. Kang, "Transformer Text Classification Model for Arabic Dialects That Utilizes Inductive Transfer," Mathematics, vol. 11, no. 24, p. 4960, 2023.

[23] S. Wu, X. Song, and Z. Feng, "MECT: Multi-metadata embedding based cross-transformer for Chinese named entity recognition," ArXiv Prepr. ArXiv210705418, 2021.

[24] T.-I. Tsai, L. Luck, D. Jefferies, and L. Wilkes, "Challenges in adapting a survey: ensuring cross-cultural equivalence," Nurse Res., vol. 31, no. 2, 2023.