IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 June 2024, accepted 11 July 2024, date of publication 22 July 2024, date of current version 31 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3431933

RESEARCH ARTICLE

DCNN: Deep Convolutional Neural Network With XAI for Efficient Detection of Specific Language Impairment in Children

KHAN MD HASIB[1], M. F. MRIDHA[2], (Senior Member, IEEE), MD HUMAION KABIR MEHEDI[3], (Member, IEEE), KAZI OMAR FARUK[3], RABEYA KHATUN MUNA[3], SHAHRIAR IQBAL[3], MD RASHEDUL ISLAM[4,5], (Senior Member, IEEE), AND YUTAKA WATANOBE[6], (Member, IEEE)

[1]Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka 1216, Bangladesh
[2]Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh
[3]Department of Computer Science and Engineering, BRAC University, Dhaka 1212, Bangladesh
[4]Department of Computer Science and Engineering, University of Asia Pacific, Dhaka 1205, Bangladesh
[5]Department of RD, Chowagiken Corporation, Sapporo 001-0021, Japan
[6]Department of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-8580, Japan

Corresponding author: Md Rashedul Islam (rashed.cse@gmail.com)

ABSTRACT Assessing children for specific language impairment (SLI) or other communication impairments can be challenging for doctors due to the extensive battery of tests and examinations required. Artificial intelligence and computer-aided diagnostics have aided medical professionals in conducting rapid, reliable assessments of children's neurodevelopmental conditions concerning language comprehension and output. Previous research has shown differences between the vocal characteristics of typically developing (TD) children and those with SLI. This study aims to develop a natural language processing (NLP) system that can identify children's early impairments using specific conditions. Our dataset contains examples of disorders, and this study seeks to (1) demonstrate the effectiveness of several classifiers in this regard and (2) select the most effective model from the classifiers. We utilized various machine learning (ML), deep learning (DL), and transformer models to achieve our objective. Our deep convolutional neural network (DCNN) model yielded excellent results, outperforming the competition with an accuracy of 90.47%, making it the top-performing model overall. To increase the accuracy and credibility of our most likely output, we have incorporated explainable AI approaches like SHAP and LIME. These approaches aid in interpreting and explaining model predictions, considering the significance and sensitivity of the topic. Additionally, we believe that our work can contribute to developing more accessible, effective methods for diagnosing language impairments in young children.

INDEX TERMS Natural language processing, language impairment, child, DCNN, XAI, LIME, SHAP.

I. INTRODUCTION

The ability to talk to and comprehend other individuals is crucial to being human. The ability to put one's thoughts and feelings into words is a reliable indicator of one's level of maturity. Any kind of language impairment or malfunction raises that risk. This emphasizes the need to identify linguistic issues as early as possible. Children mature at various speeds. Some people may take longer than others to respond or initiate conversation. A child's development follows a fairly regular pattern, with most skills appearing between the ages of 12 and 18 months and reactions appearing sometime after 6 months. When it comes to learning and reaction time, the converse is true: a child who is behind may be having difficulty. When a parent's world comes crashing down because their child can not talk

The associate editor coordinating the review of this manuscript and approving it for publication was Yu-Da Lin.

or respond normally, it is natural for them to worry. But alas, this is not always the case. Another youngster does not respond to the call of their name. Although this may not indicate an issue if other senses function well, it is still worth watching. People with language impairment have trouble learning via various means, making communicating difficult. Not smiling or playing with others (0-3 months), not mumbling (4)-7 months), making very few sounds or using no gestures (7)-12 months), not understanding what others say (7)-24 months), not being able to put two words together (18-24 months), and having trouble talking and playing with others are all signs that a child may have a language disorder (2 - 3 years). Those are some signs your kid may be exhibiting if they are dealing with communication difficulties, which may be precursors to or early indicators of conditions such as hearing loss, autism, intellectual impairment, and many more.

According to the diagnostic and statistical manual for mental disorders, all language-related disorders are identified as language disorders [1]. So, early discovery of language issues is essential to recognize as it impacts a child's development. If the children have any difficulties with language and have been diagnosed early, it may lead to intervention, then the chances become higher for improvement [2]. Many approaches have been used to detect the problem, and most are time consuming. In our paper, we have decided to use natural language processing to study examining language that is a sign of disorder. NLP is a broad field that combines computer science, linguistics, and artificial intelligence, aiming to analyze speech or language using computational methods automatically. It is a valuable technique for identifying language characteristics using potential computational methods that can automatically calculate and measure expressive languages to detect language impairments in children of early ages approximately 0 to 6 years. Our goal in this work is to implement a natural language processor on a dataset comprised of child-related information gathered from a variety of sources and symptoms, using automatic language recognition technology and deep learning methods to determine whether or not a child has a language disorder and, if so, to provide an early diagnosis that can aid both the child and their family in planning for appropriate treatment. Implications for the area are substantial because of the increased interest in using deep learning for NLP. It's popularity in NLP may be ascribed to the fact that it beats traditional machine learning models in several key areas, including accuracy, speed, automation of text analytic tasks and NLP features, and the ability to pick up on subtle nuances with relative ease [3]. In contrast to traditional machine learning models, which rely on ad hoc rules for feature extraction, this one adopts an intuitive approach. Traditional machine learning models are more intuitive when translating sources from one language to another than deep learning models. However, when using deep learning techniques, the computer learns to map the input directly depending on the output [4]. In contrast to traditional machine learning methods, deep learning is well-suited to dealing with non-linearity. It benefits human understanding of input data and judgment based on findings and provides excellent interoperability throughout training. Deep learning is helpful, but more is needed to prove the model's accuracy and efficacy. That is why we have included Explainable AI in our research: to make sure people can understand the final model.

Explainable Artificial Intelligence, often known as XAI, is a sub-field of AI still in its early stages of development. It draws on techniques from machine learning, statistics, and cognitive science. XAI is a term that refers to the tactics and methods that are utilized in the use of artificial intelligence technology to ensure that human specialists can understand the results of the solution. XAI, invented by Kuppa et al. [5], is an excellent method to illustrate the models' behavior to give a correct output. The primary purpose of XAI is to describe in depth how models generate predictions. It seeks to develop artificially intelligent systems that can be explained to humans rather than depending on high-level rules [6]. A good explanation may increase faith in the model. In general, models built using ML and deep learning approaches are referred to as black box models. They do not explain how they make a prediction based on data. If the algorithm is to make data-driven judgments, it must demonstrate that these conclusions are rational and not affected by outside factors. According to the paper [7], XAI needs to be unified with natural language processor models for understanding text classification models correctly due to the rapid growth of NLP for classifying text in different sectors to explain and understand the low-level and high-level features. It will help understand how the model works and also helps to reduce error by narrating the strength and weaknesses of the models, which will help eliminate the biases that can appear while training data. Because our study is on a sensitive issue connected to medical science, a thorough explanation of the model's prediction is required. In such a problem where the model needs to identify a disorder, depending on the model's prediction might be a major flaw. Explainable AI plays a significant role in our research, providing reliability and proper justification for the prediction. Overall, we can deduce that the approaches used in the research study improve the detection model's accuracy and effectiveness in detecting problems at an early stage and raise trust and confidence in the model.

The study's overarching goal is to create a diagnostic tool for early detection of language disorders in children. Problems communicating may be one of the earliest warning signs of health issues in youngsters. Several models, such as logistic regression (LR), decision tree (DT), shallow neural network (SNN), deep neural network (DNN), deep convolutional neural network (DCNN), Sentence fine-tuning (SeFit) for a Few-Short learning (FSL) and bidirectional encoder representations from transformers (BERT), have utilized throughout the experiment. However, the external neural network (SNN) does relatively well on the dataset

when identifying language disability. To identify language difficulties in children between the ages of 0 and 6, we combined a natural language processor with XAI in this research. To assess the model's efficacy and inspire confidence among its end-users, we also highlighted distinctive characteristics that were most useful for each choice. Following is a synopsis of the study's most significant findings:

- In this research, a dataset has been created suitable for the model, collected from a range of ages, around 0-6 years old children of various locations and organizations, which are approved by few expertise later and then annotated carefully for model implementation.
- Introduced a deep learning model known as the deep convolutional neural network (DCNN) to diagnose language impairment in children. To explain the performance of our model, we used trustworthy assessment measures such as accuracy, loss, precision, recall, f1-score, and ROC-curve.
- Deep learning models with complex structure have vindicated their data driven decision with accuracy and precision, which is difficult to interpret. However, to ensure the reliability of the model, explainable AI techniques like LIME and SHAP explanation have been used to evaluate the model's consistency, correctness, and each feature's contribution to the model by illuminating functions and logic of decisions.

The rest of the paper is organized as follows: Section II represents the related work. Section III introduces the workflow and describes a detailed explanation of methodology which is used to develop the detection of language disorder in early age using text based data. The performance analysis of the developed model evaluated using performance metrics and graph with explainable AI methods are reported in section IV. Finally, the paper is concluded, including the future implications and limitation of the work in section V and summary in Section VI.

II. RELATED WORK

Despite the importance of language and literacy throughout a person's life, there has been surprisingly little study of individuals with developmental language disorder (DLD) and their language, literacy, and cognitive abilities. Nonetheless, there are a few major exceptions. Using machine learning and network science, Borovsky et al. built prediction models to identify toddlers who would have low linguistic (LL) ability. To fine-tune the effectiveness of the approach, the authors of this piece explored parental report assessments of early linguistic competence. Several network science methods were assessed using the MacArthur-Bates Communicative Development Inventory (MBCDI) data. Infants and toddlers between the ages of 16 and 36 months old are often surveyed for this data. In this study, the author used two longitudinal datasets, namely EIRLI and LASER [8]. The researchers combined the two datasets into a younger (EIRLI 16 month and LASER 18 month) and an older (EIRLI 28 month and LASER 27 month) dataset with equivalent demographic characteristics, vocabulary size, grammatical complexity, combining words, and structural variables. Vocabulary size was referred to as the MBCDI percentile, while vocabulary structure or lexico-semantic measures included Mean Path Length, Global Clustering Coefficient, Mean Degree, Betweenness Centrality, and Harmonic Centrality. Finally, demographic variables were introduced such as education of the parents, income of the household, race, gender, and speech or language disorder history of the family. Using the nested cross validation method and the random forest model, seven of the best features from this collection of 14 were chosen for each dataset. The random forest model was used because it can provide reliable classifications of complicated datasets with many feature types (e.g., binary, categorical, numeric) and distributions. The constructed model produced strong and dependable results for the prediction of subsequent LL, with classification accuracy in individual datasets above 90% [8]. Due to variations in outcome ages and diagnostic measures, generalization performance across various datasets was limited. Less accurate predictions were made for situations in the alternate dataset.

Preterm delivery is linked to low academic attainment, poor social, emotional, and behavioral functioning, and unemployment, and may cause lifelong linguistic difficulties. By combining perinatal clinical data with the properties of diffusion MRI (dMRI), Valavani et al. sought to build a machine learning model that could reliably predict normal vs. delayed language outcomes at Corrected Gestational Age (CGA) of 2 years [9]. The authors hypothesized, supported by evidence from other research in the same field, that a combination of clinical, environmental, and imaging parameters obtained from DTI (Diffusion Tensor Imaging) might enhance the prediction of language outcomes at CGA of 2 years after premature birth. As a consequence of this, the variables that make up the proposed model are comprised of prenatal features, perinatal characteristics, postnatal characteristics, demographic characteristics, variables obtained from DTI histograms, and BAYLEY-III language assessment scores. The findings for this research came from an analysis of 89 preterm neonates who had had dMRI testing as well as a language assessment using Bayley-III at a CGA of 2 years. The clinical cut-off of 85 on the Bayley-III language assessment created a dichotomous result by dividing children into two groups. Children with scores below 85 had moderate-to-severe language impairment, while those with scores over 85 were in the normal range or above. The dataset needed some data balancing and SMOTE was found to be the best fit for the dataset. The authors investigate the similarities and differences between the Boruta, ReliefF-expRank, and random forest (RF) variable significance feature selection methodologies. The final collection of features for each feature selection procedure was determined with the help of leave-one-out cross validation. When applied to the RF classifier, a subset of eight features that were chosen using the Boruta feature selection approach provides the highest level of accuracy while also maintaining the greatest degree of

balance. The selected feature subset consists of PSFA, PSRD, and PSAD (all of which were generated from dMRI), as well as twin status (yes or no), prenatal steroid exposure (full or incomplete course), any antenatal steroid exposure (yes), and sex (male or female). The model achieved a satisfactory level of accuracy by achieving 91 percent while also achieving 86 percent sensitivity and 96 percent specificity. Finally, the researchers repeated the investigation to evaluate the performance of the model when it was given with just clinical or MRI features, which led to a decline in the performance of the model. The longitudinal cohort of preterm infants that was comprehensively phenotype with brain imaging and biological data was this study's biggest asset.

Numerous machine learning and deep learning models have been used in recent years to predict children with specific language impairment (SLI) utilizing a significant amount of effort. SLI is a language disability characterized by difficulties with speaking, comprehending spoken language, etc. The authors [1], [10], [11] utilized the LANNA children's corpus as their work's database. Since 2016, this database has been accessible via LANNA (Laboratory for Artificial Neural Network Applications) at the Czech Technical University in Prague. Yogesh Sharma and Bikesh Kumar Singh attempted to shorten the lengthy process, which is done through behavioral analysis and age-appropriate language assessments, by identifying children with SLI. The objective of their research was to characterize SLI in children using LPC (Linear Predictive Coding Coefficients) characteristics collected from their speech utterance. LPC is capable of modeling and predicting speech sequences based on its historical data. As it effectively monitors the envelope of the speech signal, it might serve as a crucial instrument for describing the quality of speech. For the classification challenge, the researchers used two supervised learning classifiers, namely naive bayes (NB) and support vector machine (SVM). The primary obstacle of this research was to retain classification accuracy while decreasing the amount of input features. However, only the top 10 and top 20 significant features were used to compare the accuracy rates of the classifiers. With a 5-fold cross-validation technique, the NB classifier with the top 20 LPC features achieved the highest accuracy of 97.9%. The speech samples were taken from children diagnosed with SLI and healthy toddlers. The utterance was used to construct the normalized log power spectrograms (LPS), the primary focus of their research. Calculated LPS for voice signals can be viewed as images.

The computed LPS is then used for training the ResNet-18 classifier. This approach proved efficient and successful in applications involving picture recognition and categorization. The researchers trained the LPS on different classifiers, including FFNN (Feed-forward neural network), ELM (Extreme-Learning Machine), and SOM (Self-organized map), however ResNet-18 was determined to be the most effective of these models. Using ResNet-18, the researchers obtained an accuracy of 99.48% and a computational cost of 0.01 GMAC [11], demonstrating that this model does not need a significant amount of processing power for the given dataset. Presently, early identification of Mild Cognitive Impairment (MCI) caused by Alzheimer's disease (AD) and Alzheimer's-type dementia (AD type dementia) is difficult. Therefore, Orimaye et al. [12] suggested an automated diagnosis approach that employs a form of deep neural network language models (DNNLM) to analyze the verbal utterances of afflicted.

Lifelong difficulties in social interaction and communication are hallmarks of the neurodevelopmental disease known as autism spectrum disorder. Autism spectrum disorders (ASD) are developmental disorders that first appear in infancy and continue throughout adulthood. For the purpose of predicting and analyzing ASD symptoms in children, adolescents, and adults, Raj and Masood looked at the viability of employing naive bayes, support vector machine, logistic regression, KNN, neural networks, and convolutional neural networks [13]. The best way to tell if a child has a language disorder is to use a natural language processor with an explainable artificial intelligence on a dataset that was compiled from various sources and the symptoms of children [14], we concluded after learning about the problem and analyzing the related research works in depth. Automatic language recognition software and deep learning techniques will be used to make this happen. Due to the autonomous nature of feature extraction in deep learning models, we were able to eschew the use of traditional machine learning approaches, which are often based on the creation of bespoke rules in order to improve performance. The model's precision has been verified using explainable AI methods like LIME and SHAP.

III. METHODOLOGY

A typical algorithmic workflow of our research has been demonstrated on figure 1. After the collection of data, they are fetched into feature extraction methods. We have implemented many deep learning models on the feature extracted data for text classification and compared among them. At the same time, we have chosen the best predicted model between the deep learning techniques. Finally, we introduced the best black box model with explainable AI techniques to make it more human understandable.

A. DATA COLLECTION

Our study is focused on automatic voice recognition for the purpose of diagnosing language disorders; hence, our dataset consists of written documents. As a whole, our study dataset comprises of inquiries, assertive speech, and responses to all three. Young children are being asked these questions and taught these stories. Our focus is on children aged 0 to 6 years. We have polled numerous children within this age range with the permission of their families, including infants from our own family, extended family, neighbors, non-governmental organizations (NGOs) who work with children, hospitals, and many more. The surveys focused mostly on

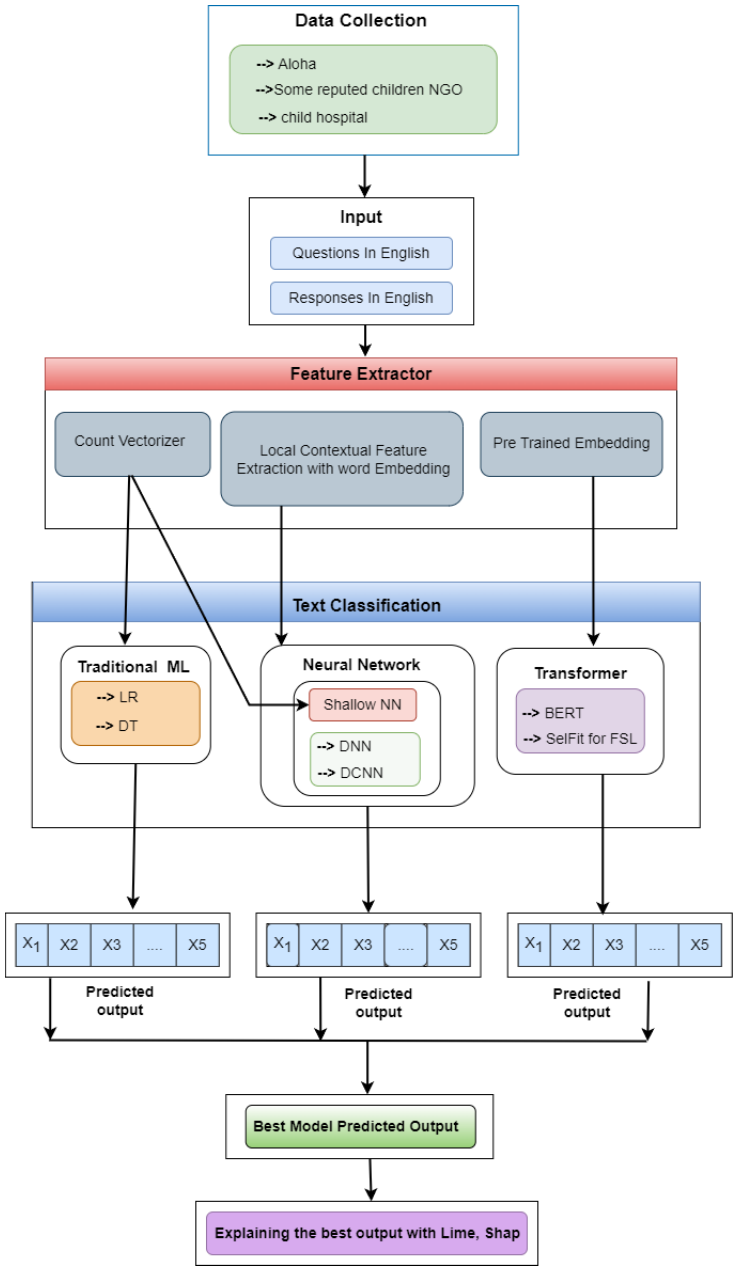

FIGURE 1. Complete process flow of the study.

eliciting responses from the children, whether via direct questioning or age-appropriate aggressive language. Any action or expression of approval in response to the question or statement counts as an answer. Children of varying ages reach several stages in their development of language. A baby who is 7-12 months old may utilize babbling consonant-vowel combinations and consonant sounds; a baby who is 12-20 months old may use gestures, identify their own name, etc.

A one-year-old should not be subjected to the same level of questioning or forceful language as a six-year-old. So, not every kid gets asked the same questions or has the same line of dialogue. However, it proved difficult to acquire information from kids of this age. It was not simple to talk to the newborns, especially since their behavior might vary depending on where you put them. Keeping an eye on the kids is a delicate and time-consuming task. It may be exceedingly challenging to collect data from children by monitoring them or conducting surveys, as these approaches are often ineffective. Unfortunately, this was the only viable option for collecting reliable information about the attitudes and actions of the students involved. The study's primary objective is to identify children with language impairments by analyzing their responses to posed questions or spoken dialogue.

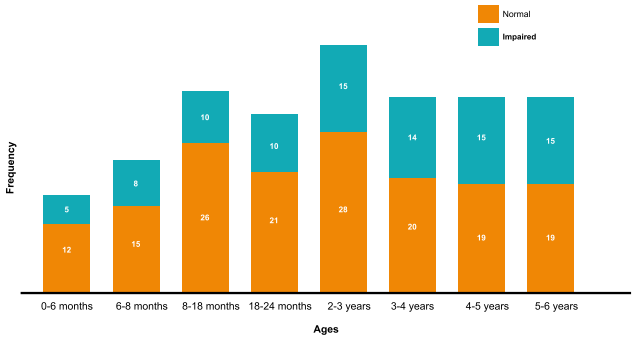

FIGURE 2. Data distribution on ages over normal and impaired children.

Since we collected the data in Bangladesh, all surveys and investigations were conducted in Bangla. All the questions and speeches prompted to the toddlers, as well as their responses, were in Bangla. All of these data are translated into English to facilitate the model's execution. For better clarity, some of our dataset examples are shown in Table 1. In the ''Comment'' columns of the table ''Check for the next step'' means the positive class and ''Normal'' means negative class of the dataset.

The responses of the youngsters to the inquiries made or the words said are crucial to the study's purpose. Responses may be anything from a simple nod to a full-on conversation, from tears to laughter to an optimistic outlook to nothing. According to responses to the questions and narration, the proposed model is effective. We also solicited the aid of various organizations that focus on infant language impairment in addition to our survey of youngsters. Many of the details we used pertain to children aged 4 to 6, and we gleaned most of this information from ALOHA Mental Arithmetic, a division of ALOHA Bangladesh (www.alohabdonline.com). The ALOHA MENTAL MATHEMATICS system [15] is a complete approach to improving one's cognitive abilities that has led to very remarkable results. A child-centered autonomous educational institution called ALOHA BANGLADESH is responsible for introducing a novel mathematics approach to the country of Bangladesh. They started experiencing substantial losses during the outbreak and in the years after the pandemic ended. Their losses may be attributed to the substantial number of students who chose not to continue their education because of a lack of technology skills and difficulty in adjusting among young pupils. Children (aged 4-13) can benefit from this training program, which has been given worldwide approval. We also confirmed the validity of the organization's survey questions to the kids. For privacy reasons, we are not include a number of other groups that address comparable concerns for children.

B. DATA ANALYSIS

We fabricated our own data set to use in this investigation. Prior to the experiment, we required to do an analysis of the dataset. Age, original Bangla speech, English translation, original Bangla answer, English translation, and label are the six columns that make up our dataset. We gathered information from kids as young as one month old and as old as sixty (60) months old. 2 to 3 years old's make up the bulk of our data set, as shown in figure 2: age, whereas infants and toddlers make up the smallest subset (ages 0 to 6 months). In addition, we can see that the number of normal samples is 160 where the number of impaired samples is 92 (out of 252 total samples) in figure 2. Despite the obvious unfairness of this situation, we refrained from using any sample methods because language disability is such a touchy subject. In this experiment, then, we have relied on actual data. Therefore, we have combined the English translations of the speech and the English translations of the responses because the age columns had no bearing on our research. Classification labels have been written in the comment fields; ''normal'' is represented by the number 0, and ''impaired'' by the number 1.

C. DATA PREPROCESSING

Data preprocessing [16] is a data mining technique used to turn raw data into a format that is convenient and easy to use. Within the context of this study, we have applied text tokenization, the removal of punctuation, the removal of superfluous special characters, and converted them to lower case.

- **Text Tokenization:** Tokenization [17] is the process of breaking down a string of text into smaller chunks. Tokens in this context might be words, letters, or even parts of words. For this experiment, we have used space to separate words in our text.
- **Remove Punctuation:** Textual data in its raw form may include extraneous elements such as HTML codes, punctuation, and special characters [18]. Here, we have utilized regular expressions to get rid of all the extraneous material. The dataset has been cleaned up by excluding punctuation and symbols that were included for clarity, such as [., ’’’|’()!?]. The substitution of every symbol with a space allowed us to break the sentence down into its component words.
- **Lower Case:** Since a computer interprets lower case and upper case letters in a text differently, it is easier for a machine to comprehend the meaning of the words when the text is presented in the same case throughout. For instance, the computer interprets the term ''Rice'' quite differently from the word ''rice''. In order to prevent these sorts of issues, the text will need to be formatted using the same case throughout, and the lower case is the one that is recommended.
- **Removal of Superfluous Special Characters:** Stopwords are the words that appear the most often in a text yet do not contribute anything of value to the meaning of the text. They are present in every kind of material, although they have no such impact on the content itself.

TABLE 1. Sample of the dataset.

Age	Speech in Bangla	Speech Translated in English	Response in Bangla	Response Translated in English	Comment
6-8 months	Eta ke re?	Who is this baby?	Smile die hat naracce	Smile with moving hand	Normal
8-18 months	cholo kheli!!	Let's play!!	No response	No response	Check for the next step
2-3 years	amake khelna ti dao !!	Give me that toy !!	Positive Response	Positive Response	Normal
3-4 years	Tumi Ki Ekhon Khabey?	Do you want to eat now?	No Response	No Response	Check for the next step
4- 5 years	Ki Koro?	What are you doing ?	Kituna	nothing	Normal

Only articles and prepositions were taken out in order to complete this task.

D. SPLIT THE DATASET INTO TRAINING AND TESTING SETS

In this section,the dataset has been divided into a training set and a test set. 75% of the data in the dataset was utilized for training, while the remaining 25% was used for testing. Among 75% of train data, 10% data was kept for validation to validate our model in the training process.

E. MODEL IMPLEMENTATION

As we have said earlier, our dataset is very small for training heavy machine learning [19] and deep learning [20] models, but we are trying to show a direction on how the language impairment problem of children can be solved. To demonstrate their performance on our dataset, we implemented two ML (logistic regression, decision tree) models with count vectorizer, three DL (shallow neural network, deep neural network, deep convolutional neural network) models with both count vectorizer and embedding layer, and two transformers (bidirectional encoder representations from transformers [21] and Sentence fine-tuning for FSL models.

- **Logistic Regression (LR):** Logistic regression [22] is a kind of supervised learning approach for categorization that is used to create predictions on the probability of an outcome variable. In most cases, LR refers to binary logistic regression, which has binary target variables. However, there are two more kinds of target variables that may be predicted using multinomial logistic regression (MLR). There are adjustments for categorization purposes. If n cases with m features have k classes, the m matrix points towards component B being computed. The probability for class j with the exception of the class, is as in the below Eqn.:1:

$$P_j(X_i) = \frac{\exp(X_i B_j)}{\sum_{j=1}^{k*1} \exp(X_i B_j) + 1} \tag{1}$$

- **Decision Tree (DT):** In the field of machine learning, the DT [23] method is classified as a supervised learning technique. As opposed to other supervised learning algorithms, the decision tree technique can be applied to both regression and classification problems. DT is used to build a training model that can predict the class or value of the target variable based on a few basic rules learned from historical data (training data).

$$\text{Gini} = \sum_{i \neq j} p(i)p(j) \tag{2}$$

$$H(S) = \sum_{c \in C} -p(c)\log_2 p(c) \tag{3}$$

$$IG(A, S) = H(S) - \sum_{t \in T} p(t)H(t) \sum_{t \in T} p(t)H(t) \tag{4}$$

When using DT to forecast a record's class label, we begin at the top of the tree. The values of the root property are compared to those of the corresponding fields in the record. The comparison then leads us along the branch that leads to the next node. DT uses entropy theory to classify instances. Entropy is a metric used in information theory to gauge how pure or uncertain a set of observations is. It controls the way a decision tree decides how to divide data. The above equation 2 is the *gini* impurity of entropy, Eqn. 3 is the equation of entropy($H(S)$) and Eqn. 4 is the equation of information gain(IG). Here, Information gain(IG) is measured as the difference between the entire dataset entropy and the splitting attribute entropy. the algorithm divides the data into two or more sets using information gain and entropy, splitting is accomplished using the most relevant qualities to produce classes as separate as feasible. Entropy, as stated in the formula, checks the impurity of the result class in a subset with the properties of p in any S dataset.

- **Shallow Neural Network (SNN):** The term "neural network" automatically implies that the system in question must contain a large amount of information that is hidden from view. However, there is a type of neural network known as SNN [24] that is made up of only a few of these layers.

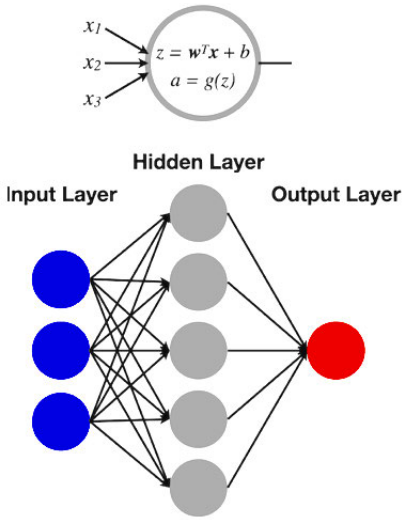

FIGURE 3. Shallow neural network (SNN) with one hidden layer.

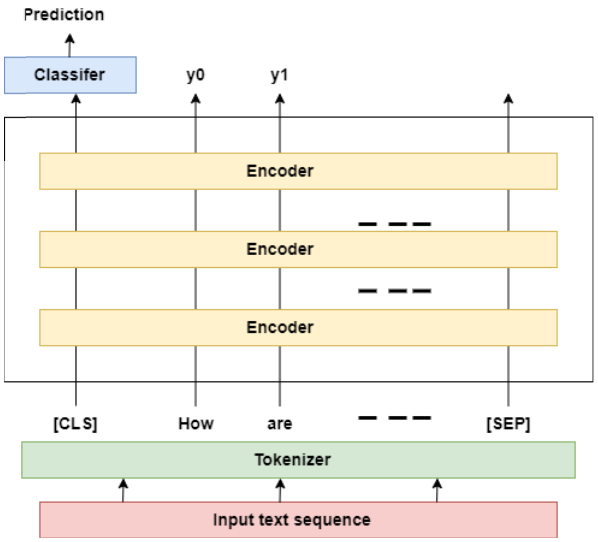

FIGURE 4. The working procedure of BERT.

TABLE 2. Parameter table of deep neural network (DNN).

Layer (Type)	Output Shape	Param #
embedding_21 (Embedding)	(None, 100, 50)	5600
flatten_1 (Flatten)	(None, 5000)	0
dense_1 (Dense)	(None, 100)	500100
dense_2 (Dense)	(None, 100)	10100
dense_3 (Dense)	(None, 100)	10100
dense_4 (Dense)	(None, 100)	10100
dense_5 (Dense)	(None, 100)	10100
dense_6 (Dense)	(None, 1)	101
Total params: 546,201		
Trainable params: 546,201		
Non-trainable params: 0		

In shallow neural networks, the number of hidden layers is either one or two. Occasionally, there may be no hidden layers at all. In the work that we have done, we have used two hidden layers. There are 10 neurons in the first layer that have a relu activation function, and there is just one neuron in the very top layer that has a sigmoid activation function. Both the count vectorizer and the embedding layer of the SNN have been the subject of research and development [25]. Figure 3 represents the shallow neural network architecture.

- **Deep Neural Network (DNN):** Deep neural network [26] is an extended version of a shallow neural network. The main difference between them is that SNN has one to two hidden layers, but DNN can have more than one hidden layer between the input and output layer. In our study, five hidden layers with an embedding layer are used. Table 2 depicts the parameter of the deep neural network (DNN).
- **Bidirectional Encoder Representations from Transformers (BERT):**
The BERT [27] framework is one that is open-source and used for NLP. BERT employs context to assist computers in correctly interpreting content that may be unclear. The BERT framework was first trained on the content of Wikipedia and may be further refined with the use of question and answer datasets. The architecture of our BERT model is depicted in figure 4.
Transformers are a kind of deep learning model that are used in BERT. This model ensures that every output element is connected to every input element, and that weightings are dynamically calculated based on how these elements are related to one another. Language models could either read text sequentially from left to right or right to left, but not both at the same time. BERT understands both interpretations. This was made feasible by the development of Transformers, and the resulting property is referred to as bidirectionality. In contrast to transformers, BERT only requires an encoder unit, and the decoder component, as the name indicates and as seen in figure 4, will be removed. Each encoder is made up of the same layers as its corresponding transformer, namely Self-Attention and Feed Forward Neural Networks.
- **Sentence fine-tuning (SeFit) for Few-Short learning:**
Traditionally, building machine learning systems meant collecting a lot of data and using it to train machine learning algorithms. It is costly to collect, categorize, and validate massive amounts of data. In many situations, we do not have access to large data sets and must rely on a small number of examples to make decisions.
Few-shot learning is a trending technique [28] in machine learning, in which a model generates predictions based on a small number of training samples.

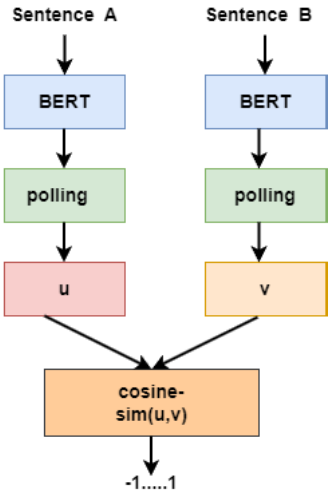

FIGURE 5. Sentence fine-tuning (SeFit) for Few-Short learning.

Sentence Transformer (ST) is a widely-used technique for semantic search, semantic similarity, and grouping or clustering. The concept behind the sentence transformer is encoding a unique vector representation of a phrase based on its semantic signature. During contrastive training, a transformer based model is adapted into a siamese architecture and used to build the representation. The figure 5 of a siamese network with BERT attempts to decrease the gap between semantically similar statements and increase the distance between semantically different ones.

Sentence Transformers produce highly effective representations when used to large-scale comparisons between sentence-pairs, which is a prevalent scenario in information retrieval tasks. In this task, we have utilized this technique for effective classification of language impairment in children.

- **Proposed Deep Convolutional Neural Network (DCNN):**

In figure 6 shows how our model work in this dataset. In artificial intelligence (AI), researchers used CNN [29] to extract features from images. However, nowadays, it is used for text classification in NLP. It has two types layers which are convolutional and pooling layers. The initial stage of a convolutional network is the convolutional layer.

In computer vision CNN preserves the 2D spatial orientation of an image. Like images, texts also have an orientation. Texts have a one-dimensional structure where word order matters as opposed to being two-dimensional. Here all words of the training example are represented as n-dimensional vectors. Here the filters of the convolutional neural network helps in extracting specific features from input n-dimensional vector. After extracting the features the pooling layers helps to reduce the dimension by selecting the maximum, minimum or average element from the region of the feature map covered by the filter. The following equation 5 is the equation of convolution operation between input and kernel.

$$y_j = \sum_{c=0}^{n_c-1} \sum_{k=-p}^{p} x_{c,j-k} w_{c,k} \tag{5}$$

The following equation 6 is the equation of input gradient.

$$\frac{\partial \mathcal{L}}{\partial x_{c,i}} = \sum_{k=-p}^{p} \frac{\partial \mathcal{L}}{\partial y_{i+k}} w_{c,k} \tag{6}$$

And the equation 7 is the equation of parameter gradient of DCNN.

$$\frac{\partial \mathcal{L}}{\partial w_{c,k}} = \sum_{j=0}^{m-1} \frac{\partial \mathcal{L}}{\partial y_j} x_{c,j-k} \tag{7}$$

In our proposed deep convolutional neural network we have applied few fully connected (FC) layers to fit more non linear pattern in the data. In our proposed DCNN model, we have used five 1D Conv layers with 'relu' activation function, starting with an embedding layer demonstrated in table 3. After conv layers, we have used one 1D GlobalMaxPooling layer. Finally, there are a total of five FC layers. The first four FC layers each have 100 neurons, and the final layer has 1 neuron with a sigmoid activation function [30] since we are doing binary classification.

TABLE 3. Parameter table of deep convolutional neural network (DCNN).

Layer (Type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 100, 100)	11200
conv1d (Conv1D)	(None, 96, 128)	64128
conv1d_1 (Conv1D)	(None, 92, 128)	82048
conv1d_2 (Conv1D)	(None, 88, 128)	82048
conv1d_3 (Conv1D)	(None, 84, 128)	82048
conv1d_4 (Conv1D)	(None, 80, 128)	82048
global_max_pooling1d	(None, 128)	0
dense_9 (Dense)	(None, 100)	12900
dense_10 (Dense)	(None, 100)	10100
dense_11 (Dense)	(None, 100)	10100
dense_12 (Dense)	(None, 100)	10100
dense_13 (Dense)	(None, 1)	101
Total params: 446,821		
Trainable params: 446,821		
Non-trainable params: 0		

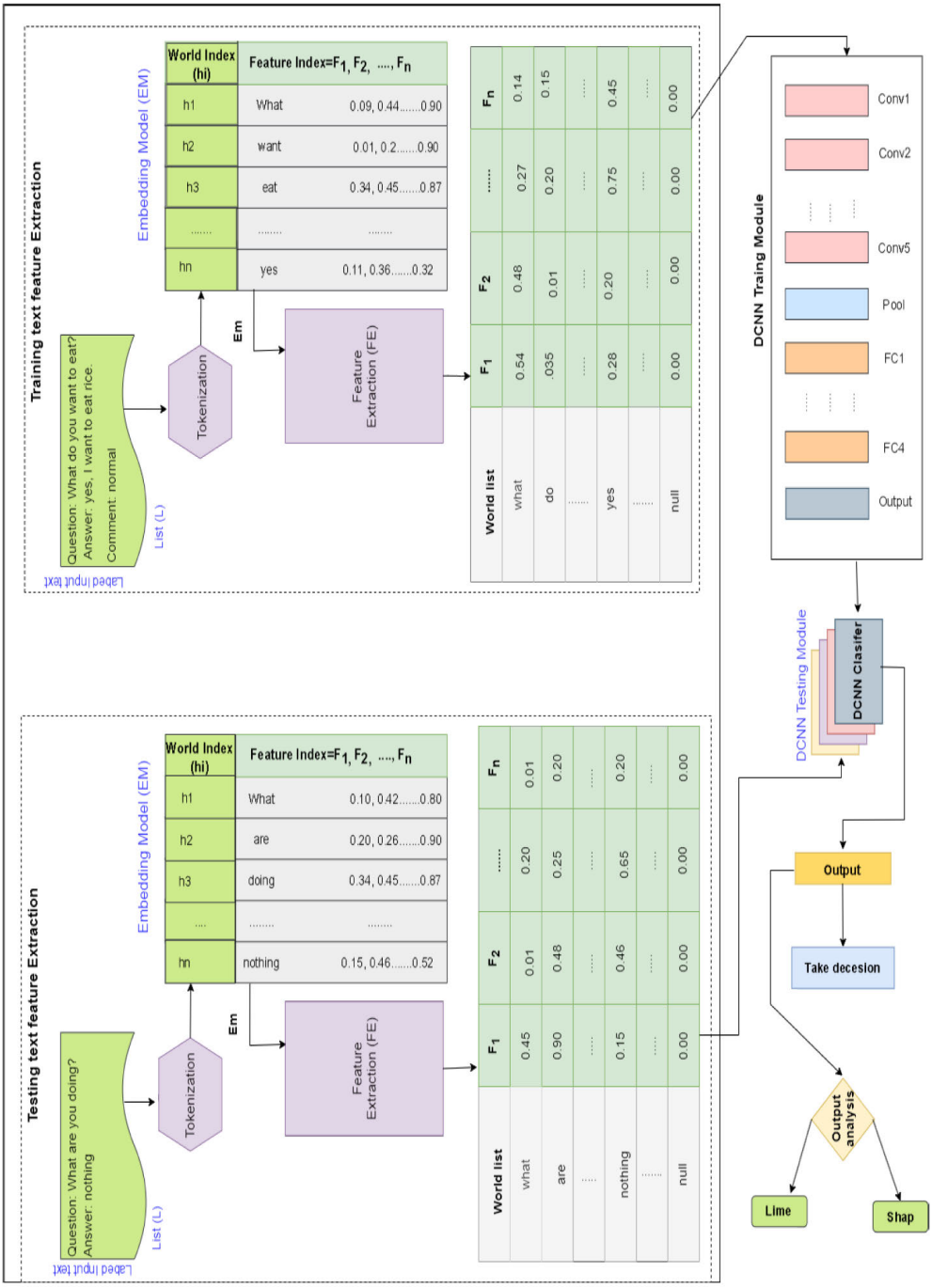

FIGURE 6. Proposed deep convolutional neural network (DCNN) with embedding.

F. EXPLAINABLE AI (XAI)

Artificial intelligence has grown rapidly in recent years. AI models have begun to exceed human intelligence at a rate no one could have expected. As models get more exact and precise, it is tougher to explain their complicated mathematical calculations. Mathematical abstraction does not help users trust a model's choices. Explainable AI refers to approaches that explain an AI model's decision-making process [31]. This new field of AI has huge promise, with increasingly advanced approaches each year. SHAP, DeepSHAP, DeepLIFT, CXplain, and LIME are popular XAI approaches.

- **LIME:** In the field of artificial intelligence, LIME [32] stands for Local Interpretable Model Agnostic Explanations. The purpose is to teach artificial intelligence systems to comprehend forecasts made by humans. As a kind of "local explanation," it works best when talking about specific cases. With the help of the supplied data,

LIME may [33] create fictitious information with just some of the true properties. For example, while working with textual data, many copies of the original are created, each with a different number of randomly selected letters. Next, the newly created false information is segmented into several groups (classified). So, we can see the impact that certain keywords have on the text's overall categorization.

- **SHAP:** For each feature, SHAP [34] calculates its Shapley value, which represents how much weight that feature has in the prediction as a whole. In our research article, we employ additive shapley explanations, as well as locally interpretable and model-agnostic shapley explanations. The final result (prediction) is the sum of these factors plus the baseline (average prediction throughout the validation set; a value closer to 1.0 indicates a higher likelihood of being a fake). SHAP also allows us to use color-coded violin plots derived from all predictions to quickly and intuitively highlight the relevance of a feature, and use these plots to establish a correlation between low/high feature values and an increase/decrease in output values.

IV. RESULTS ANALYSIS

We have reported the outcomes of our experiments and the findings of our performance analyses in this part. Initially, we investigated the classifiers' ability to foretell the results of language impairment in children using assessment measures. Finally, we assess how well the recommended model, DCNN works. The efficacy of the model is evaluated using a training vs. validation accuracy, loss graph, and confusion matrix. Finally, we compare our findings to those of other investigations.

A. EVALUATION MEASURES

Our customized model has been judged using the standards established by the scikit-learn package [35]. The primary goal is to pick the model that fits our needs the best. We have experimented with a variety of parameters, including batch size, learning rate (lr), epochs, activation function, loss function, and the number of dense layers; however, the settings that gave us the best results were batch size $= 5$, lr $= 0.001$, epochs $= 20$, and 5 dense layers with a ''sigmoid'' activation function and a ''binary_crossentropy'' loss function. We have taken advantage of the early stopping mechanism in order to prevent overfitting. In order to make meaningful comparisons across algorithms, we need to assess metrics such as accuracy, precision, recall, and f1-score. True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are used as proxies for accuracy, precision, recall, and f-score in the following equation 8.

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + FP + FN + TN} \\ \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \\ F1 * \text{ score} &= \frac{2 \text{ precision recall}}{\text{precision} + \text{recall}} \end{aligned} \tag{8}$$

The following paragraphs elaborate on each of the aforementioned steps:

- True Positive (TP) is the total number of occurrences that have been properly labeled as having a positive value or yes (1) by the created model M* after the labeled instances have been updated.
- True Negative (TN) refers to the total number of instances that were properly identified by the created model as having a negative value or the value zero.
- False Positive (FP) is the total amount of occurrences classified incorrectly, which means that the machine predicts the value as positive/yes (1) but its actual value is negative/no (0) by the generated model M* after updating the labeled instances. In other words, FP is the total amount of occurrences that were incorrectly classified.
- A computer forecasts the value as negative/no (0), but its real value is positive/yes (1) as determined by the created model. This is what is meant by the term ''false negative,'' which refers to the total quantity of occurrences categorized wrongly.
- Accuracy: The accuracy of the test is determined by the percentage of the total data that is correctly classified.
- Precision is a statistic that is used to evaluate how precise a class is in comparison to the actual world.
- Recall: A recall is a kind of measure that is used to evaluate how well prepared a class is.
- F-Score: The F-score was designed in order to take into consideration the possibility of both false positive and false negative results.

Table 4 and 5 present the comparisons of various models' performance to analyze which model works well for the non-linear text-based dataset to detect language disorders among a certain range of children. Several evaluation metrics, including accuracy, f1-score, recall, and precision [36], have been used to assess the models. We have implemented three types of models in our research paper: the traditional machine learning model, transformers, and deep neural networks.

Considering the experiment of different models, the deep learning models work better than traditional machine learning models and transformer models for our research. Among the deep learning models, the results of DCNN achieved the highest goal in terms of the evaluation metrics which shows the outcome of the model is good for the research. Table 4 represents the performance of training set in terms of accuracy and loss while train the whole model. According to the result, the performance of DCNN is the best in terms of accuracy and loss for training set where the model obtained highest accuracy 95.77% among the models, with loss 14.11% which is lesser than the second highest accuracy

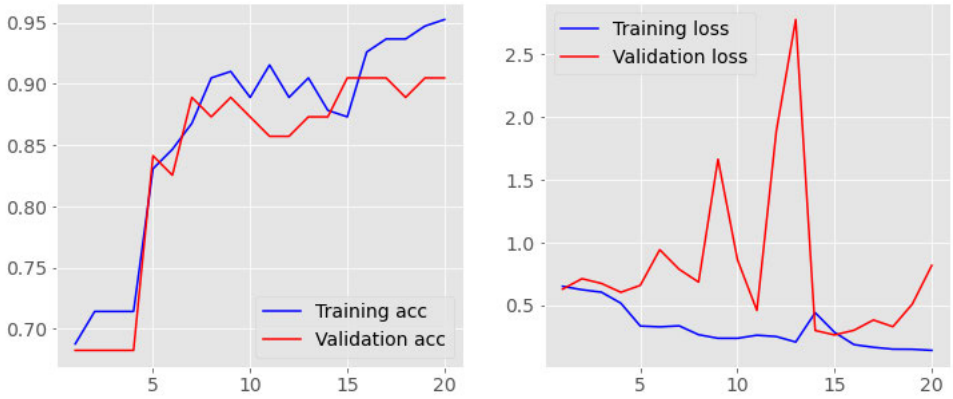

FIGURE 7. Training vs validation accuracy and loss graph for our proposed deep convolutional neural network (DCNN).

TABLE 4. Training accuracy and loss comparison between different models.

Model	Accuracy	Loss
SNN (Count vectorizer)	86.24	34.94
SNN (embedding)	93.12	13.30
DNN (embedding and 4 hidden layers)	94.71	15.70
BERT	72.49	56.71
Proposed DCNN	**95.77**	**14.11**

TABLE 5. Test accuracy (A), precision (P), recall (R) and F1 score (F1) comparison between different models.

Model	A	P	R	F1
LR	84.12	69.56	84.21	76.19
DT	80.26	68.08	64.00	72.73
SNN (Count vectorizer)	82.54	**90.91**	50.00	64.51
SNN (embedding)	87.30	80.00	80.00	80.00
DNN (embedding& hidden layers)	88.88	76.00	95.00	84.44
SeFit for FSL	60.31	36.36	42.10	39.24
BERT	60.31	45.65	**99.00**	62.68
Proposed DCNN	**90.47**	81.81	90.00	**85.71**

model DNN. On the other hand, BERT performs very poor with lowest accuracy of 72.49% and highest loss 56.71%.

Table 5 represents the results of test set to show the actual performance of models where the traditional machine learning models: logical regression and decision tree have been obtained 84.12% and 80.26% accuracy. Though is greater than transformer models: SeFit for FSL and BERT. On the other hand, the deep learning models: shallow neural network, deep neural network, and DCNN perform better than traditional machine learning model and transformer

TABLE 6. Highest, Avg. and lowest train and test accuracy of the proposed DCNN model.

	Highest accuracy	Average accuracy	Lowest accuracy
Train	95.77	95.19	94.87
Test	90.47	90.15	89.61

models, where shallow neural network with count vectorizer, shallow neural, DNN and DCNN embedded accuracy are 82.54%, 87.30%, 88.88% and 90.47% respectively. This indicates DCNN achieved the goal of the best outcome in terms of performance metrics. In the same way, DCNN obtained the highest score in terms of F1 Score with 85.72%. On the other hand, in terms of recall and precision, BERT and SNN obtained the highest score with 99% and 90.91% respectively. Thus, we can say even though, there is no particular model that obtained highest score in each performance metric. However, the performance of DCNN model is better in general than any other models and the transformer models, especially SeFit for FSL performance, are very poor in terms of accuracy, precision and f1 score among all since our dataset is small.

We developed our model using trail and error method. After getting a poor result, every time we fine tuned the model. Every time the model shows a different result, before finalizing the result, we run the model more than 10 times and pick the one with the highest result. In table 6 shows the highest, average and lowest train and test accuracy.

Figure 7 represents the loss and accuracy charts for our proposed deep convolutional neural network during training and validation (DCNN).

Quantitative analysis has been done based on the confusion matrix of the models performed on the text-based dataset to evaluate the performance of four best models that performs well for detection. In the figure 8(a) and 8(b) the confusion matrix of models logistic regression and decision tree, then BERT 8(c), SeFit for FSL 8(d), shallow neural network 9(a), embedded shallow neural network 9(b),

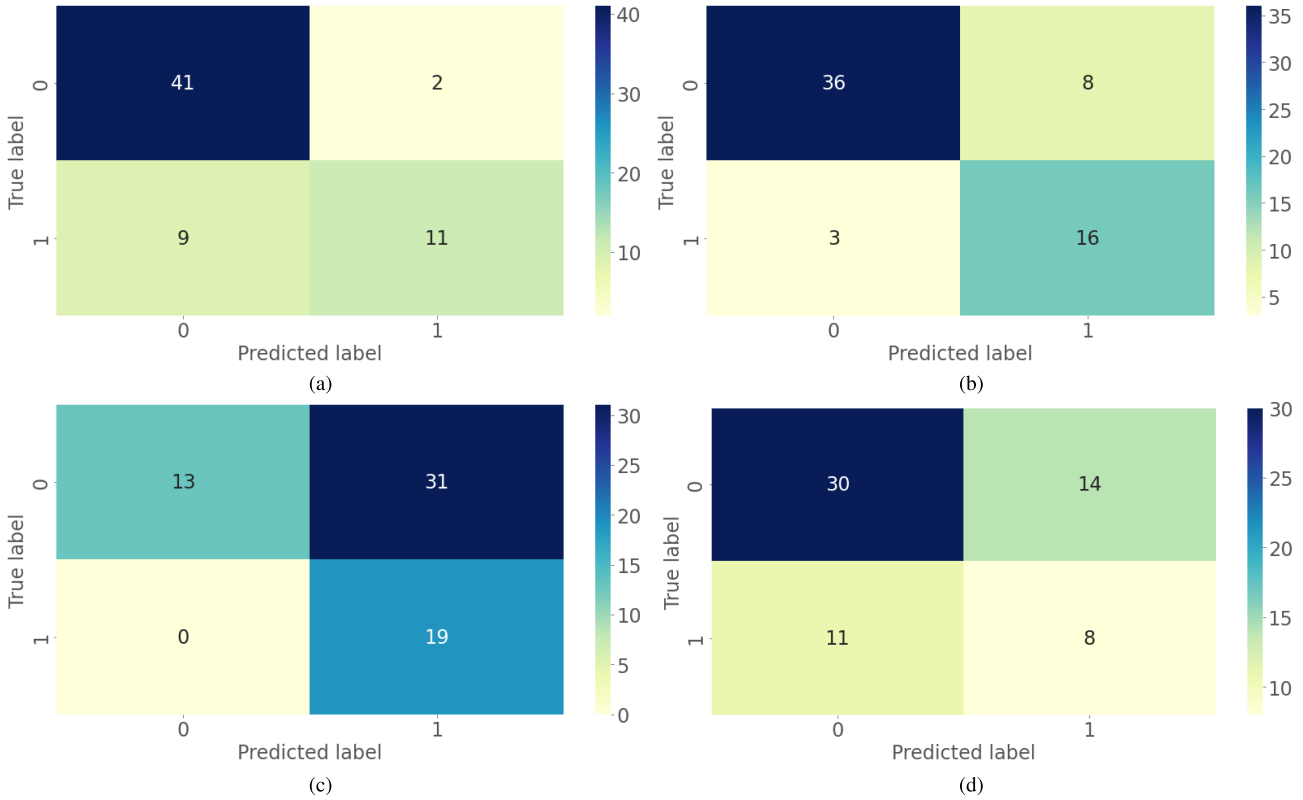

FIGURE 8. Confusion matrix for (a)LR (b)DT (c)BERT (d)SeFit for FSL. Here, 0 = normal, 1 = check for next step/ impaired.

DNN 9(c) and DCNN 9(d) have been shown the normal response of mis-classified samples are 2, 8, 31, 14, 1, 5, 6 and 4 respectively. On the other hand, in the case of abnormal or impaired response, the samples are mis-classified by the models 9, 3, 0, 11, 8, 1, 1 and 2 respectively. From the analysis, we come to know that, Logical Regression model gives the highest and SNN gives the lowest misleading or False Negative results among eight of the models for normal response. In the mean while, the transformer model SeFit gives the highest and SNN with embedding and DNN give the lowest misleading results for the abnormal or impaired response.

In figure 10 represents the ROC curve with 0.90 AUC (area under the curve) of the DCNN model. Here, AUC 0.90 means the model can identify the positive and negative classes with 90% chance of being correct.

From our experiments, we assert that the performance of deep neural network based models performs better than pre-trained and fine tuned models and traditional machine learning for our dataset, and among the deep neural network models, DCNN performs best on our dataset in terms of confusion metrics, accuracy, F1-score, precision, and area under the curve among our experimented models. Apart from the DCNN, other models like- DNN, SNN, logistic regression and decision tree perform well. In contrary, the transformer based methods BERT and SeFit do not perform well in our relatively small dataset as they require a huge number of training data for better prediction.

B. RESULT ANALYSIS USING EXPLAINABLE AI

LIME stands for Local Interpretable Model Agnostic Explanations which is a XAI tool that gives interpretation of models by providing locally faithful explanation to reflect the behavior of the classifier. In our research, we have decided to exhibit the prediction for each test from the dataset for each example to make the model transparent and human understandable enough which may help to enhance the models performance for classifying the instance more accurately by modifying the alerting feature that has effect on prediction [37]. Among the four models, LIME has been implemented on DNN model due to it is better performance in model experiment, shown in result analysis. Notice that for each class, in figure 11, 12, 13 and 14 the words right side on the line are positive and left side of the line are negative. For instance, in figure 11, 'calling by name' with no response is positive is one of the sign of impairment for more than 6 months old children with 0.77 prediction probability. In figure 12, the answer 'that thing' for the question 'what do you want' is the positive response for normal children with range more than 1 years old children. The probability of identifying language disorder is 0.92 for the answer of

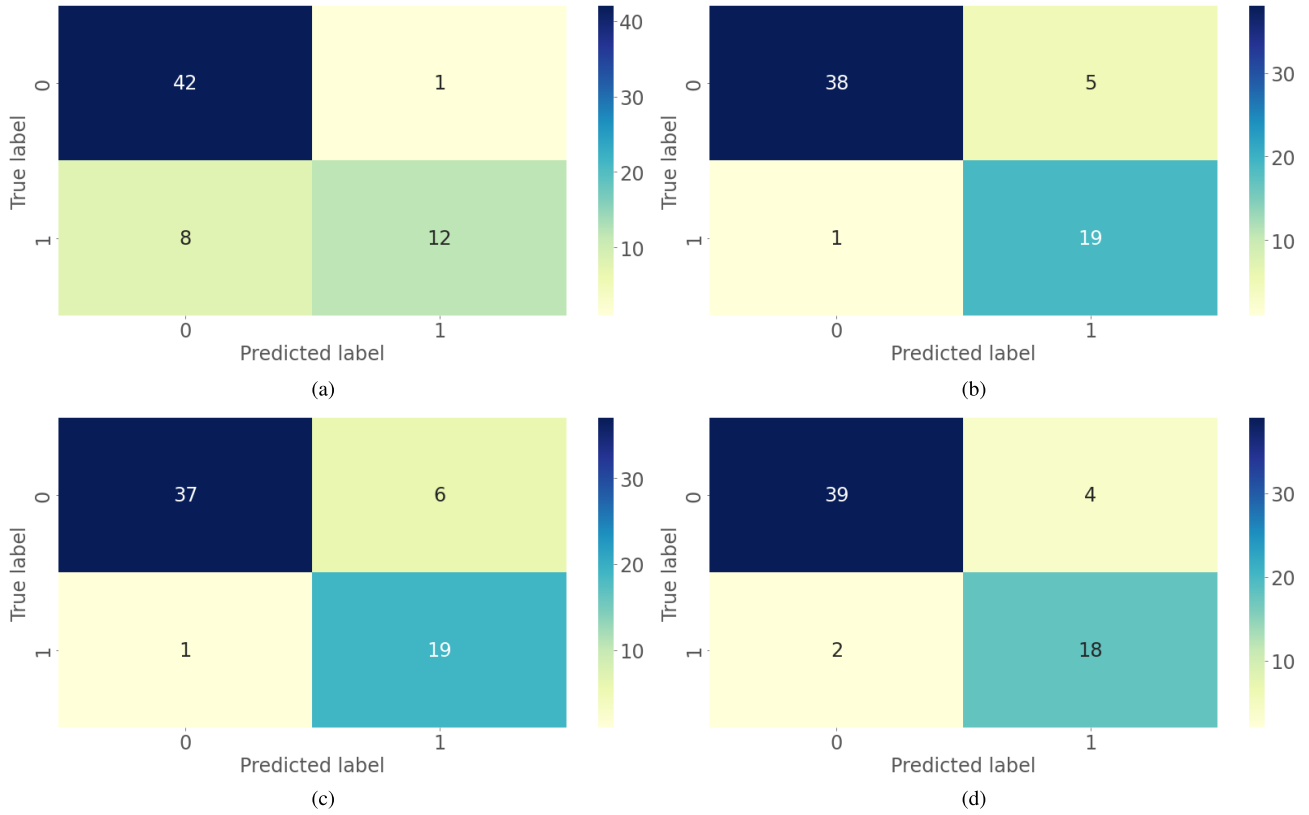

FIGURE 9. Confusion matrix for (a) SNN with count vector (b) SNN with embedding (c) DNN (d) DCNN. Here, 0=normal, 1=check for next step/ impaired.

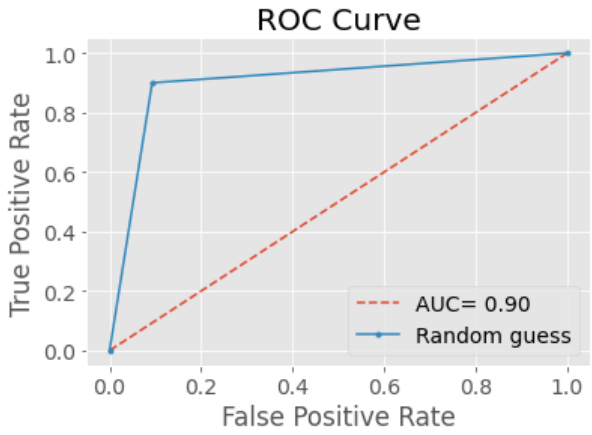

FIGURE 10. ROC curve for proposed DCNN model.

'smile' of the question 'do you want rice' is the positive response for more than 2 years old children in figure 13. In the last figure 14, the question 'where is your mama?' has the 0.69 smiling answer for the Normal response for more than 1 years old children.

SHAP attempts to explicate the prediction of an observation by calculating the contribution of each feature to the prediction. It displays the list of significant features, from the most significant to the least significant, as well as the feature that contributes the most to a certain prediction in the dataset. Each feature contributes a SHAP value to the prediction.

The Y axis in the graph from figure 15 depicts the features, while the X axis reflects their values in relation to the predictions. The features that have a substantial influence on the dataset's prediction are shown in figure 15. Since 'eat' has the largest value on the x axis, it is evident that it has the most impact on the predictions. Similarly, ''no'' is the second most important feature in the predictions. And 'what' has the least influence on the predictions shown in the preceding graph.

The figure 16 describes the SHAP values for a specific piece of data from our collection. The aggregate SHAP scores might indicate the positive or negative contribution of each feature to the prediction. The pink shade represents toddlers with positive impairments, whereas the blue color represents toddlers with negative impairments. The question of the above data was ''Do you want to eat rice?'' and there was no response to the question. Looking at the scale of figure 16, we can see that the feature ''reaction'' has the most influence on this prediction being positively hindered, along with the features ''rice'', ''eat'', ''to'', and ''want'', which are all colored pink. And the features ''you'' and ''no'' have degraded values and are tinted blue, but they do not provide reliable predictions. Consequently, we might assert that the child is positively impaired.

Similar to the preceding description, the above figure 17 depicts a second observation of the prediction of a certain

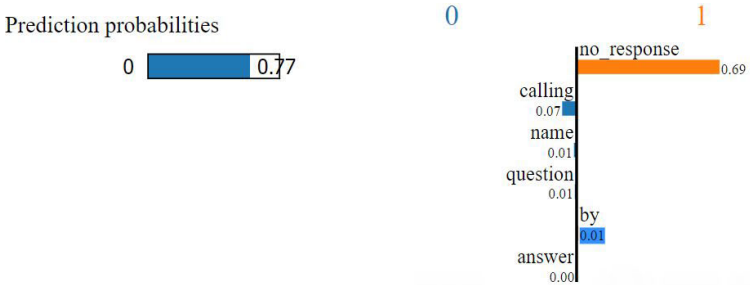

FIGURE 11. Explainable AI result for input data "question: calling by name? answer: no response".

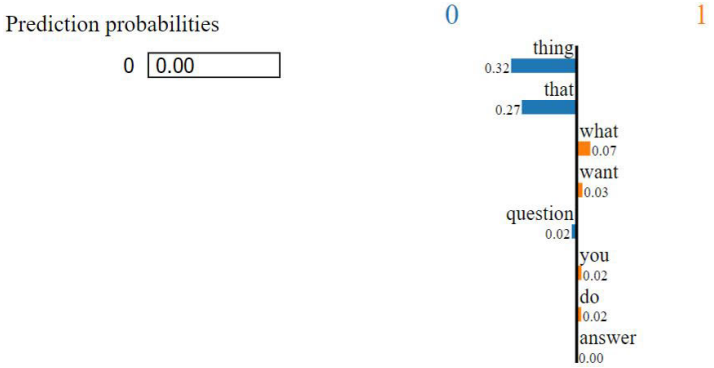

FIGURE 12. Explainable AI result for input data "question: what do you want? answer:that thing".

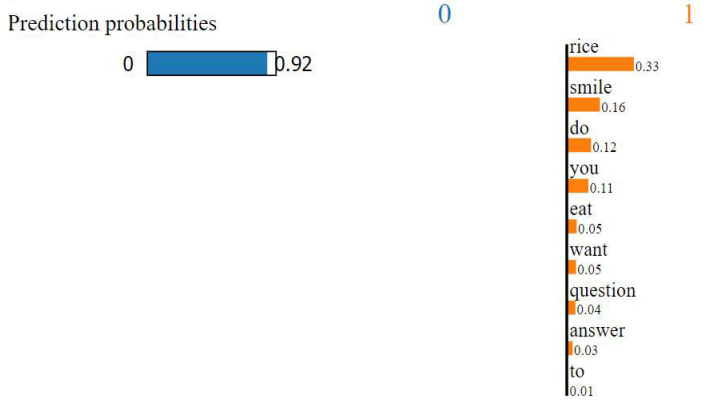

FIGURE 13. Explainable AI result for input data "question:do you want to eat rice? answer: smile".

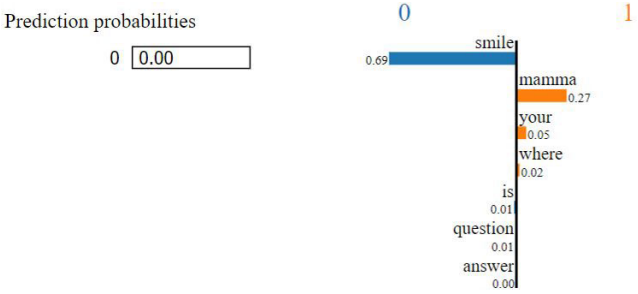

FIGURE 14. Explainable AI result for input data "question: where is your 'mamma'? answer: smile".

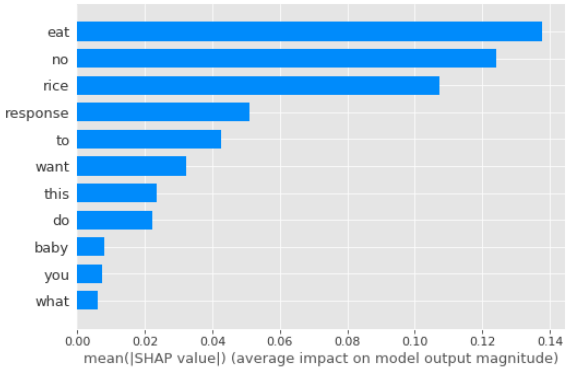

FIGURE 15. Impact of features on model's output predictions.

data, where the question is identical to the first. This time, the response to the inquiry was "smile". As the "smile" feature has a very little percentage of pink shade on the scale and the majority of the other features have blue sections, we can conclude that the toddler is negatively impaired.

FIGURE 16. Visualizing features influence on models prediction using SHAP for a random sample (1).

FIGURE 17. Visualizing features influence on models prediction using SHAP for a random sample (2).

FIGURE 18. Visualizing features influence on models prediction using SHAP for a random sample (3).

FIGURE 19. Visualizing features influence on models prediction using SHAP for a random sample (4).

As in earlier instances, we can similarly explain the figure 18 shown above. Only ''rice'' has a negligible contribution to positive impairment, whereas other features contribute to negative impairment. ''I'' and ''want'' contribute the most to the negatively impacted sections.

In similar fashion, the above figure 19 shows another explanation of a prediction. It indicates that ''no'' and ''respond'' account for the bulk amount of portion and contribute to the positively impaired. Moreover, there are no such features making any contribution towards negative impairment. Thus, it is evident that the aforementioned prediction of figure 19 is a positive impairment.

C. COMPARISON OF ACCURACY WITH SOME EXISTING LITERATURE

To facilitate comprehension, Table 7 compares the aforementioned studies with respect to the characteristics used, datasets analyzed, and precision with which the results were obtained. The higher performance of our model over the other random forest classifier led us to choose the DCNN proposed model for the comparison. Though our dataset is small in size, which is why other work surpasses us with good accuracy.

V. LIMITATIONS AND FUTURE WORK

A. LIMITATIONS

Although every effort was made to achieve the primary objective of the work, namely the development of a complete diagnosis tool, there were several impediments that complicated the process. These include:

- NLP requires a large vocabulary of words, syntax, and grammar of a language, however, there is a limited amount of data available for children's language. As a result, it was difficult to prepare a generalized diagnostic tool.
- Dealing with children's language is especially challenging as there is a wide variation in language acquisition capability in children. For instance, some children learn to speak in later phase of childhood, which is often considered normal and not indicative of any particular disorder. So the program will need further analysis to identify patterns that are not regular but still be considered normal.

B. SCOPE OF FUTURE WORK

As mentioned above, some areas remain that can be improved or modified to reflect a better outcome. These can be summarized as follows:

- Data volume and variation can be increased to improve the accuracy of the results. In particular, subjects with known language acquisition problems and disorders can be included in a larger scope so that the diagnosis capacity of the code can be benchmarked with real cases.
- Medical and psychological information from medical doctors and psychologists can be included to improve

TABLE 7. Comparison table with some existing literature.

Reference	Dataset	Feature	Best Classifier	Accuracy
[8]	EIRLI & LASER	lexico-semantic feature, demographic variables	Random Forest	90%
[1]	LANNA children corpus	LPC (Linear Predictive Coding Coefficients)	Naïve Bayes	97.9%
[10]	LANNA children corpus	glottal source and MFCC (Mel-frequency cepstral co-efficient)	FFNN (feed-forward neural network)	98.82%
[11]	LANNA children corpus	log power spectrograms (LPS), calculated from speech utterance and can be viewed as image	ResNet-18	99.48%
[14]	ASD Screening Data from UCI repository	Demographic variables (age, sex, nationality, screening score)	CNN	98.3%
[38]	Saarbrücken Voice Disorders (SVD)	MFCC, MFCC-glottal	SVM+wav2vec 2.0	62.77% M,55.36% W
Our best classifier	**Our Dataset**	**Questions & responses from the child.**	**DCNN**	**90.47%**

the analysis criteria. Doctors generally utilize strict professional guidelines to diagnose a child with a developmental disorder. Although such guidelines are difficult to translate into objective parametric analysis, efforts can be made to correlate medical analysis to the impairment.

- The data can be collected through a device that will be with a child for 24 hours continuously. Then we can have a huge amount of data for every single class. Hence, advanced transformer methods can be implemented to enhance the accuracy of the detection program.

VI. CONCLUSION

Specific Language Impairment (SLI) is a communication issue that hinders the acquisition and development of language abilities in children who do not exhibit any auditory impairments. Specific Language Impairment has the potential to impact a child's oral communication skills, auditory comprehension abilities, reading proficiency, and written expression. SLI, sometimes referred to as developmental language disorder, language delay, or developmental dysphasia, is a recognized condition.

This study uses artificial intelligence to analyze data from children under six to detect language-based illnesses including hearing impairment and speech delay. Unfortunately, human communication is unreliable. In this field, much research has been done to find a solution. However, we were disappointed to find no language technique for identifying the issue among Bengali-speaking youngsters. This sickness might be expressive or receptive, and this project aims to build a natural language processing method for reliable medical diagnosis.

Two machine learning models with count vectorizers (logistic regression and decision tree), three deep learning models (shallow neural network, deep neural network, and deep convolutional neural network), and two transformer models sentence fine-tuning on Few-Short learning and transformer bidirectional encoder representations were used. The DCNN model has the highest accuracy and f1 score.

DATA AVAILABILITY STATEMENT

The datasets for this study can be found in the link: https://github.com/hkabirmehedi/Children-Language-Impaired-Dataset

REFERENCES

[1] Y. Sharma and B. K. Singh, "Prediction of specific language impairment in children using speech linear predictive coding coefficients," in *Proc. 1st Int. Conf. Power, Control Comput. Technol.*, Jan. 2020, pp. 305–310.

[2] B. Pandey, D. Kumar Pandey, B. Pratap Mishra, and W. Rhmann, "A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5083–5099, Sep. 2022.

[3] P. K. Athira, C. J. Sruthi, and A. Lijiya, "A signer independent sign language recognition with co-articulation elimination from live videos: An Indian scenario," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 3, pp. 771–781, Mar. 2022.

[4] S. Khan, M. Fazil, V. K. Sejwal, M. A. Alshara, R. M. Alotaibi, A. Kamal, and A. R. Baig, "BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4335–4344, Jul. 2022.

[5] A. Kuppa and N.-A. Le-Khac, "Black box attacks on explainable artificial intelligence(XAI) methods in cyber security," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[6] H. Hagras, "Toward human-understandable, explainable AI," *Computer*, vol. 51, no. 9, pp. 28–36, Sep. 2018.

[7] Z. Li, X. Wang, W. Yang, J. Wu, Z. Zhang, Z. Liu, M. Sun, H. Zhang, and S. Liu, "A unified understanding of deep NLP models for text classification," *IEEE Trans. Vis. Comput. Graphics*, early access, Dec. 4, 2022, doi: 10.1109/TVCG.2022.3184186.

[8] A. Borovsky, D. Thal, and L. B. Leonard, "Moving towards accurate and early prediction of language delay with network science and machine learning approaches," *Sci. Rep.*, vol. 11, no. 1, pp. 1–4, Apr. 2021.

[9] E. Valavani, M. Blesa, P. Galdi, G. Sullivan, B. Dean, H. Cruickshank, M. Sitko-Rudnicka, M. E. Bastin, R. F. M. Chin, D. J. MacIntyre, S. Fletcher-Watson, J. P. Boardman, and A. Tsanas, "Language function following preterm birth: Prediction using machine learning," *Pediatric Res.*, vol. 92, no. 2, pp. 480–489, Aug. 2022, doi: 10.1038/s41390-021-01779-x.

[10] M. K. Reddy, P. Alku, and K. S. Rao, "Detection of specific language impairment in children using glottal source features," *IEEE Access*, vol. 8, pp. 15273–15279, 2020.

[11] K. Kotarba and M. Kotarba, "Efficient detection of specific language impairment in children using resnet classifier," *Signal Process., Algorithms, Archit., Arrangements, Appl.*, vol. 1, pp. 169–173, Nov. 2020.

[12] S. O. Orimaye, J. S.-M. Wong, and C. P. Wong, "Deep language space neural network for classifying mild cognitive impairment and Alzheimer-type dementia," *PLoS One*, vol. 13, no. 11, Nov. 2018, Art. no. e0205636.

[13] L. S. Baron, A. Gul, and Y. Arbel, "With or without feedback?—How the presence of feedback affects processing in children with developmental language disorder," *Brain Sci.*, vol. 13, no. 9, p. 1263, Aug. 2023.

[14] S. Raj and S. Masood, "Analysis and detection of autism spectrum disorder using machine learning techniques," *Proc. Comput. Sci.*, vol. 167, pp. 994–1004, Oct. 2020.

[15] *ALOHA Mental Arithmetic: A Comprehensive Path to Mental Excellence*. Accessed: Jul. 23, 2024. [Online]. Available: https://www.alohabdonline.com/

[16] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text preprocessing for text mining in organizational research: Review and recommendations," *Organizational Res. Methods*, vol. 25, no. 1, pp. 114–146, Jan. 2022, doi: 10.1177/1094428120971683.

[17] J. J. Webster and C. Kit, "Tokenization as the initial phase in NLP," in *Proc. 14th Conf. Comput. Linguistics (COLING)*, vol. 4. USA: Association for Computational Linguistics, 1992, pp. 1106–1110.
[18] P. Durga and D. Godavarthi, "Deep-sentiment: An effective deep sentiment analysis using a decision-based recurrent neural network (D-RNN)," *IEEE Access*, vol. 11, pp. 108433–108447, 2023.
[19] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–5, May 2021.
[20] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep learning-based text classification: A comprehensive review," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–40, Apr. 2022.
[21] S. Yu, J. Su, and D. Luo, "Improving BERT-based text classification with auxiliary sentence and domain knowledge," *IEEE Access*, vol. 7, pp. 176600–176612, 2019.
[22] J. Kapusta, D. Držík, K. Šteflovic, and K. S. Nagy, "Text data augmentation techniques for word embeddings in fake news classification," *IEEE Access*, vol. 12, pp. 31538–31550, 2024.
[23] G. Sharma, X.-P. Zhang, K. Umapathy, and S. Krishnan, "Audio texture and age-wise analysis of disordered speech in children having specific language impairment," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102471.
[24] H. Cai, Z. Li, C. Yan, J. Liu, and A. Yin, "A shallow neural network based short text classifier for medical community question answering system," in *Proc. IEEE 8th Annu. Int. Conf. CYBER Technol. Autom., Control, Intell. Syst.*, Jul. 2018, pp. 1537–1541.
[25] A. Kaspi et al., "Genetic aetiologies for childhood speech disorder: Novel pathways co-expressed during brain development," *Mol. Psychiatry*, vol. 28, no. 4, pp. 1647–1663, Sep. 2022.
[26] J. Wang, Y. Li, J. Shan, J. Bao, C. Zong, and L. Zhao, "Large-scale text classification using scope-based convolutional neural network: A deep learning approach," *IEEE Access*, vol. 7, pp. 171548–171558, 2019.
[27] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-dependent sentiment classification with BERT," *IEEE Access*, vol. 7, pp. 154290–154299, 2019.
[28] V. Clay, G. Pipa, K.-U. Kuhnberger, and P. König, "Development of few-shot learning capabilities in artificial neural networks when learning through self-supervised interaction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 1, pp. 209–219, Jan. 2024.
[29] A. J. Syed, D. J. Durrani, N. Shahid, W. Khan, and A. Muhammad, "Expression detection of autistic children using CNN algorithm," in *Proc. Global Conf. Wireless Opt. Technol. (GCWOT)*, Jan. 2023, pp. 1–5.
[30] H. Pratiwi, A. P. Windarto, S. Susliansyah, R. R. Aria, S. Susilowati, L. K. Rahayu, Y. Fitriani, A. Merdekawati, and I. R. Rahadjeng, "Sigmoid activation function in selecting the best model of artificial neural networks," *J. Phys., Conf. Ser.*, vol. 1471, no. 1, Feb. 2020, Art. no. 012010, doi: 10.1088/1742-6596/1471/1/012010.
[31] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
[32] M. Tulio Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?': Explaining the predictions of any classifier," 2016, *arXiv:1602.04938*.
[33] O. Moussa, M. Mostfa, and S. El-Araby, "Evaluation of the antibacterial effect of garlic with lime on streptococcus mutans in children," *Al-Azhar Dental J. Girls*, vol. 9, no. 3, pp. 525–530, Jul. 2022.
[34] A. Kumar, S. Dikshit, and V. H. C. Albuquerque, "Explainable artificial intelligence for sarcasm detection in dialogues," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–13, Jul. 2021.
[35] N. Fang, X. Fang, K. Lu, and E. Asare, "Online incremental mining based on trusted behavior interval," *IEEE Access*, vol. 9, pp. 158562–158573, 2021.
[36] R. C. Morales-Hernández, J. G. Jaguey, and D. Becerra-Alonso, "A comparison of multi-label text classification models in research articles labeled with sustainable development goals," *IEEE Access*, vol. 10, pp. 123534–123548, 2022.
[37] T. A. J. Schoonderwoerd, W. Jorritsma, M. A. Neerincx, and K. van den Bosch, "Human-centered XAI: Developing design patterns for explanations of clinical decision support systems," *Int. J. Hum.-Comput. Stud.*, vol. 154, Oct. 2021, Art. no. 102684.
[38] S. Tirronen, S. R. Kadiri, and P. Alku, "Hierarchical multi-class classification of voice disorders using self-supervised models and glottal features," *IEEE Open J. Signal Process.*, vol. 4, pp. 80–88, 2023.

KHAN MD HASIB received the B.Sc. degree from the Computer Science and Engineering Department, Ahsanullah University of Science and Technology, Dhaka, Bangladesh, in 2018, and the M.Sc. degree from the Department of Computer Science and Engineering, BRAC University, Dhaka, in 2022. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Software Engineering, The University of Western Australia, Crawley, WA, Australia. He has more than five years of teaching and four years of research experience in computer science. He was an Assistant Professor with the Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka. He has authored or co-authored over 40 research papers in highly recognized journals, book chapters, and conference proceedings. He is working on several projects, such as efficient detection of specific language impairment in children, LLM over transfer models for low resource language, autism spectrum disorder in adults from screening results with xAI, and deep learning techniques for analyzing and visualizing restaurant food reviews. His research interests include applied machine learning, natural language processing, natural language generation, representation learning, low-resource language, and interpretability.

M. F. MRIDHA (Senior Member, IEEE) received the Ph.D. degree in AI/ML from Jahangirnagar University, in 2017. He is currently an Associate Professor with the Department of Computer Science, American International University-Bangladesh (AIUB). Before that, he was an Associate Professor and the Chairperson of the Department of CSE, Bangladesh University of Business and Technology. He was a CSE Department Faculty Member with the University of Asia Pacific and the Graduate Head, from 2012 to 2019. For more than ten years, he has been with the master's and undergraduate students as a Supervisor of their thesis work. His research experience, within both academia and industry, resulted in over 120 journals and conference publications. His research work contributed to reputed journals, such as *Scientific Reports* (Nature), *Knowledge-Based Systems*, *Artificial Intelligence Review*, IEEE Access, *Sensors*, *Cancers*, and *Applied Sciences*. His research interests include artificial intelligence (AI), machine learning, deep learning, natural language processing (NLP), and big data analysis. He has served as a program committee member for several international conferences/workshops. He served as an Associate Editor for several journals, including *PLOS One* journal. He served as a Reviewer for reputed journals and IEEE conferences, such as HONET, ICIEV, ICCIT, IJCCI, ICAEE, ICCAIE, ICSIPA, SCORED, ISIEA, APACE, ICOS, ISCAIE, BEIAC, ISWTA, IC3e, ISWTA, CoAST, icIVPR, ICSCT, 3ICT, and DATA21.

MD HUMAION KABIR MEHEDI (Member, IEEE) received the B.Sc. degree in computer science from BRAC University, Dhaka, in 2022, where he is currently pursuing the M.Sc. degree in computer science and engineering. He has more than three years of research experience in advanced artificial intelligence, machine learning, and deep learning. Particularly in the areas of applied machine learning, medical image processing, computer vision, and natural language processing, he has been actively engaged in joint research activities. He has publications in prestigious book chapters and conference proceedings. Some of his journal articles are under review in reputed journals. He served as a Reviewer for reputed conferences, such as PRICAI and IJCNN.

KAZI OMAR FARUK received the B.Sc. degree in computer science and engineering from the Ahsanullah University of Science and Technology (AUST). He is currently pursuing the M.Sc. degree in computer science and engineering with BRAC University. He has more than two years of research experience in advanced artificial intelligence, machine learning, and deep learning. He has published eight research papers in highly reputed conferences within a very short time. His research interests include genetic optimization, federated learning and its application, autoencoders, collaborative filtering, multicriteria decision-making, recommendation systems, the application of natural language processing, and deep learning in recommendation systems

RABEYA KHATUN MUNA received the B.Sc. degree in computer science from BRAC University. She is currently a Software Engineer with SELISE. She has more than one year of research experience in computer science. She had one research paper published at a prestigious conference. Her research interests include machine learning, the IoT, computer vision, deep learning, and natural language processing.

SHAHRIAR IQBAL received the B.Sc. degree in computer science and engineering from BRAC University. He has about a year of research experience in machine learning, computer vision, and deep learning. He is conducting research on federated learning and its applications. He had one research paper published at a prestigious conference.

MD RASHEDUL ISLAM (Senior Member, IEEE) received the B.Sc. degree in computer science and engineering from the University of Rajshahi, Rajshahi, Bangladesh, in 2006, the M.Sc. degree in informatics from Högskolan i Borås (the University of Borås), Borås, Sweden, in 2011, and the Ph.D. degree in electrical, electronic, and computer engineering from the University of Ulsan, Ulsan, South Korea, in 2016. He was a Senior Architect with the Research and Development Department, Exvision Corporation, Tokyo, Japan; a Visiting Researcher (a Postdoctoral Researcher) with the School of Computer Science and Engineering, The University of Aizu, Japan; a Graduate Research Assistant with the Embedded System Laboratory, University of Ulsan; an Assistant Professor with the Department of Computer Science and Engineering, University of Asia Pacific (UAP), Dhaka, Bangladesh; and a Lecturer with the Department of Computer Science and Engineering, Leading University, Sylhet, Bangladesh. He is currently a Chief Researcher of computer vision and AI with Chowagiken Corporation, Japan, and an Associate Professor (on leave) with the Department of Computer Science and Engineering, UAP. Also, he has good experience in professional IT system analysis and development. His research interests include machine learning, signal and image processing, HCI, health informatics, bearing fault diagnosis, and others. He is a member of the IEEE Computer Society and the IEEE Computational Intelligence Society. He is also a PC member of several international conferences. He served as the Secretary for the Organizing Committee of the 19th International Conference on Computer and Information Technology 2017 (ICCIT 2017); the Organizing Chair for the Organizing Committee of the ACM-ICPC Dhaka Regional Site 2017; the Head of the Self-Assessment Committee (SAC) for the Department of CSE under IQAC, University of Asia Pacific; a Coordinator for the MCSE Program, Department of CSE, UAP; a Convener for the Software and Hardware Club, Department of CSE, UAP; a Coordinator for the Admission Committee, Department of CSE, UAP; and a Treasurer for Bangladesh Advanced Computing Society. He is a Reviewer of several journals, such as the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE ACCESS, *Applied Science*, *Multimedia Tools and Applications*, *Cluster Computing*, *Shock and Vibration*, *Journal of Information Processing Systems*, and others.

YUTAKA WATANOBE (Member, IEEE) received the M.S. and Ph.D. degrees from The University of Aizu, in 2004 and 2007, respectively. He was a Research Fellow with Japan Society for the Promotion of Science (JSPS), The University of Aizu, Japan, in 2007. He is currently a Senior Associate Professor with the School of Computer Science and Engineering, The University of Aizu. He is also the Director of i-SOMET. He was a Coach of four ICPC World Final teams. He is a Developer of the Aizu Online Judge (AOJ) System. His research interests include intelligent software, programming environments, smart learning, machine learning, data mining, cloud robotics, and visual languages. He is a member of IPSJ.