# A Transformer-CNN Hybrid Model for Cognitive Behavioral Therapy in Psychological Assessment and Intervention for Enhanced Diagnostic Accuracy and Treatment Efficiency

Dr. Veera Ankalu. Vuyyuru[1], G Vamsi Krishna[2], Dr.S.Suma Christal Mary[3], Dr.S.Kayalvili[4],
Abraheem Mohammed Sulayman Alsubayhay[5]

Assistant Professor, Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Vaddeswaram, 522502, A.P, India[1]
VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad[2]
Professor, Department of Information Technology, Panimalar Engineering College, Poonamalle, Chennai[3]
Associate Professor, Department: Artificial Intelligence, Kongu Engineering College, Perundurai, Tamilnadu[4]
Faculty of Computing, Universiti Teknologi Malaysia (UTM), Johour, Malaysia[5]

*Abstract*—The use of Cognitive Behavioral Therapy (CBT) as a method for psychological assessment and intervention has shown to be quite successful. However, by utilizing advancements in artificial intelligence and natural language processing techniques, the diagnostic precision and therapeutic efficacy of CBT can be significantly improved. For CBT in psychological evaluation and intervention, we suggest a unique Transformer-CNN hybrid model in this work. The hybrid model combines the strengths of the Transformer and Convolutional Neural Network (CNN) architectures. While the CNN model successfully extracts local and global features from the input sequences, the Transformer model accurately captures the contextual dependencies and semantic linkages in the text data. It intends to enhance the model's comprehension and interpretation of the complex linguistic patterns involved in psychological evaluation and intervention by merging these two algorithms. On a sizable collection of clinical text data, which includes patient narratives, treatment transcripts, and diagnostic reports, we undertake comprehensive experiments. The proposed Trans-CNN hybrid model outperformed all other methods with an impressive accuracy of 97%. In diagnosing psychiatric problems, the model shows improved diagnosis accuracy and offers more effective therapy advice. Our hybrid model's automatic real-time monitoring and feedback capabilities also make it possible for prompt intervention and customized care during therapy sessions. By giving doctors a formidable tool for precise evaluation and efficient intervention, the suggested approach has the potential to revolutionize the field of CBT and enhance patient outcomes for mental health. In order to improve the diagnostic precision and therapeutic efficacy of CBT in psychological evaluation and intervention, this work provides a transformational strategy that combines the advantages of the Transformer and CNN architectures.

*Keywords—CBT; psychological assessment; intervention; diagnostic accuracy; treatment efficiency; Transformer; CNN; NLP*

## I. INTRODUCTION

Cognitive Behavioral Therapy (CBT) is a widely used approach in psychological assessment to identify and understand various cognitive, emotional, and behavioral patterns that contribute to mental health disorders [1]. It focuses on exploring and modifying these patterns to improve psychological well-being [2]. The assessment helps identify specific areas of concern, symptoms, and the impact of these factors on the individual's daily life and functioning. Initial Assessment in CBT is defined as the therapist conducts an initial assessment to gather information about the client's history, presenting problems, and treatment goals. Structured or semi-structured interviews, questionnaires, and rating scales may be used to assess symptoms, cognitive biases, and maladaptive behavior [3]. The formulation outlines the interplay between thoughts, emotions, behaviors, and external factors, providing a conceptual framework for treatment planning. The assessment findings guide the development of a tailored treatment plan in CBT [4]. Specific treatment goals and strategies are identified to address the individual's unique difficulties and promote positive changes in thoughts, emotions, and behaviors. Assessment continues throughout the course of therapy to monitor progress, reassess symptoms, and adjust treatment strategies if needed. Regular feedback and outcome measures are used to evaluate the effectiveness of interventions and make informed treatment decisions. Psychological assessment in CBT is an ongoing process that informs treatment decisions and guides the therapeutic interventions. It helps therapists gain a comprehensive understanding of the individual's cognitive and emotional functioning, enabling them to tailor treatment approaches to meet specific needs and achieve positive outcomes.

Cognitive Behavioral Therapy (CBT) has emerged as a highly effective approach for psychological assessment and intervention. However, the diagnostic accuracy and treatment efficiency of CBT can be further enhanced by leveraging

advances in artificial intelligence and natural language processing techniques. In this study, we propose a novel Transformer-CNN hybrid model for CBT in psychological assessment and intervention. The hybrid model combines the strengths of both the Transformer and Convolutional Neural Network (CNN) architectures. The Transformer model captures the contextual dependencies and semantic relationships in the text data, while the CNN model effectively extracts local and global features from the input sequences. By integrating these two architectures, we aim to improve the model's ability to understand and interpret the nuanced language patterns associated with psychological assessment and intervention. We conduct extensive experiments on a large dataset of clinical text data, including patient narratives, therapy transcripts, and diagnostic reports. In cognitive-behavioral therapy (CBT), the classification process involves categorizing individuals based on their presenting problems, symptoms, and assessment results. Standardized diagnostic standards, such as those listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), are first applied in this procedure, to evaluate whether the individual meets the criteria for a specific mental health disorder. Diagnostic interviews, structured questionnaires, and rating scales are commonly used to gather information about the individual's symptoms, duration, and functional impairment [5].

In addition to diagnostic classification, the classification process in CBT extends to determining appropriate treatment recommendations for individuals. Based on the assessment findings, therapists match the individual's diagnosis or presenting problems with evidence-based treatment protocols. This includes specific therapeutic techniques, intervention strategies, and self-help resources tailored to the individual's needs. It's important to note that the classification process is not static and may evolve throughout the course of therapy. Ongoing evaluation and monitoring of symptoms, progress, and treatment outcomes help therapists assess the effectiveness of interventions and make adjustments as needed. The classification process may be refined based on the individual's response to treatment and changes in symptom presentation. The classification process in CBT provides a framework for understanding an individual's mental health concerns, identifying specific symptoms, and tailoring treatment recommendations [6]. By systematically categorizing individuals based on their assessment results, therapists can provide targeted interventions and support to address their unique needs and improve psychological well-being. The results demonstrate that our hybrid model achieves superior performance compared to traditional CBT methods and standalone Transformer or CNN models. The model exhibits enhanced diagnostic accuracy in identifying psychological disorders and provides more efficient treatment recommendations [7]. Furthermore, our hybrid model also enables automated real-time monitoring and feedback during therapy sessions, facilitating timely intervention and personalized treatment. We utilize a sequential process that involves data collection, data preprocessing, model design, input representation, training and optimization, and evaluation. The dataset is split into training, validation, and test sets, and appropriate loss functions and optimization algorithms are employed to train the model. Relevant parameters, such as accuracy, precision, recall, F1-score, or area under the receiver operating characteristic curve (AUC-ROC), are used to assess the model. The suggested model has the power to transform the field of CBT by providing clinicians with a powerful tool for accurate assessment and effective intervention, leading to improved mental health outcomes for patients. The primary objective of this study is to develop a novel approach to CBT in psychological evaluation and intervention that leverages the benefits of both the Transformer and CNN models. By doing so, we aim to improve the accuracy of psychiatric problem diagnosis and provide more effective therapy advice to patients. Here, introduces a Trans-CNN hybrid model that effectively merges the Transformer and CNN architectures. The CNN model excels at extracting local and global features from input sequences, while the Transformer model accurately captures contextual dependencies and semantic linkages in text data. By combining these two algorithms, our hybrid model offers a powerful tool for precise evaluation and efficient intervention in mental health care. Overall, this study presents a transformative approach that combines the strengths of the Transformer and CNN architectures to enhance the diagnostic accuracy and treatment efficiency of CBT in psychological assessment and intervention [8]. The key contributions of this research lie in the development of a hybrid model that combines Transformer and CNN architectures, aiming to enhance diagnostic accuracy, treatment efficiency, and personalization are as follows:

- Data is collected from the posts of depressed person.

- Initially, Text pre-processing methods such as is employed to pre-process the text in the collected data posts.

- Feature extraction is employed by utilizing transformer encoder using the self-attention method.

- Classification using Transformer-CNN Hybrid Model for CBT in psychological assessment and intervention.

The sections that make up the research's structure are as follows: Section II describes related works that demonstrate a comparison of the proposed method with several other methods, Section III discusses the problem statement, Section IV defines the recommended approach for the Trans-CNN method, Section V elaborates the evaluation metrics, and Section VI describes the outcome and potential future research.

## II. RELATED WORKS

The scientific community is becoming more interested in recognizing emotions as a result of the many fields it may be used in, such as health-care or traffic security schemes. D. Griol *et al.* [9] presents a speech- and facial-based multimodal emotion identification method. Researchers tested numerous transfer-learning methods, more especially embedding extraction, and fine-tuning, for the speech-based medium. The PANNs framework's CNN-14 was fine-tuned to produce the most excellent accuracy results, demonstrating that instruction was more reliable when it didn't start from scratch and when the assignments were comparable. Using saliency maps and

face pictures, we provide architecture for emotion-based facial recognizers consisting of a bi-LSTM with an attention mechanism after a pre-trained Spatial Transformer Network. Regardless of the domain change, the error analysis revealed that the frame-based systems could pose issues if employed directed for completing a video-based task. In order to resolve this discrepancy and make use of the inherent information of these pre-trained models, this brings up a new avenue of study. Finally, using a late fusion technique and these two modalities together, we classified eight emotions with 80.08% accuracy on the RAVDESS dataset using a subject-wise 5-CV assessment. The findings showed that various modalities convey pertinent data to identify users' emotional states, and their integration enhances the efficiency of the system. Despite domain adaptation, this study notes that frame-based systems may run into issues when utilized directly for video-based activities. This shows that using pre-trained models efficiently for video-based emotion identification may provide some difficulties. To overcome these issues and enhance the system's effectiveness in real-world circumstances, more study is required.

Y. Chen *et al.* [10] proposed the research paper "Screening Children's Intellectual Disabilities with Phonetic Features, Facial Phenotype and Craniofacial Variability Index,". In this study, a kind of neurological deficit condition called intellectual disability (ID) is brought on by congenital disorders or things that happen after birth. If early screening for this syndrome was effective, the condition of patients may be improved, and their capacity for self-care increase. Clinical interviews are the only method for achieving early ID screening, and they necessitate the active involvement of medical specialists and assets associated with medicine. The methods employed in this research paper comprises that the analysis of young participants' phonetic traits and facial phenotype has been suggested as a potential way for screening ID. In order to compute the craniofacial variability index (CVI) and determine the likelihood of ID, the geometric aspects of the faces of the participants and phonetic characteristics of the subjects' voices are initially determined from interview footage. Furthermore, a technique for enhanced ID screening centered on visual and phonetic traits is established using machine learning algorithms. The Results suggested that the technique was assessed using three feature sets, including phonetic, geometric, and CVI characteristics. Accuracy performed best when it was near 80%. In conclusion, based on the findings obtained with the three feature sets, it can be said that, with further development, the suggested approach may be used in a clinical environment. The diversity of the dataset used to assess the suggested strategy is not disclosed by this approach. The method's applicability to larger populations may be in doubt if the dataset is restricted in terms of characteristics or certain ID subtypes.

Z. Chen *et al.* [11] proposed the research study "Detecting Reddit Users with Depression Using a Hybrid Neural Network," In this study, a common mental health problem, depression is thought to afflict 3.8% of the world's inhabitants. Additionally, it is a significant factor in the global burden of disability. People are increasingly using social media sites like Reddit to communicate their problems and health issues, such as depression, and to ask for help from other users in online groups. By analyzing millions of messages for possible solutions, it creates a fantastic opportunity for machines to identify social media users who are depressed. Since deep learning techniques are simple to use, analyze information quickly, and produce cutting-edge results on a variety of NLP tasks, they have started to take the lead in the machine learning and natural language processing (NLP) fields. In this study, researchers present a hybrid deep learning model to identify depressed people from Reddit postings using a pre-trained sentence BERT (SBERT) and convolutional neural network (CNN). To learn the meaningful representation of semantic information in every message, the sentence BERT is employed. CNN makes it possible to further alter these embeddings and to identify user behavior trends over time. Using the Self-reported Mental Health Diagnoses (SMHD) data, we educated and assessed the model's ability to detect Reddit users who are depressed. The hybrid deep learning model exceeded the most recent results (F1 scores of 0.79 by other machine learning models in the literature) by achieving an accuracy rating of 0.86 and an F1 score of 0.86. The outcomes demonstrate the viability of using a hybrid model to detect depression in people. The hybrid model may easily be tweaked and used for other text classification tasks as well as many clinical scenarios, despite its validation to identify sadness using Reddit postings. The Self-reported Mental Health Diagnoses (SMHD) data are mentioned in this study as a training and assessment tool. Self-reported diagnoses, however, are not always precise or trustworthy, which might affect the quality of the labeled data and, in turn, the efficiency of the predictive algorithm.

S. Aleem *et al.* [12] presents ML techniques to identify the depressive disorder. In this research, Stress, worry, and today's fast-paced lives have all had a significant psychological impact on people's minds over time. The worldwide technical advancement in healthcare digitizes the relevant data, making it possible to map the many facets of the human condition more precisely than with conventional measurement methods. For analyzing the vast quantity of data in the healthcare sector, machine learning (ML) has been recognized as an effective method. In order to forecast the likelihood of mental diseases and, consequently, execute prospective therapies, ML techniques are being used in the field of psychological wellness. Numerous machine-learning methods for recognizing and diagnosing sadness are discussed in the above article. Classification, deep learning, and ensemble are the three categories of ML-based sadness detection techniques. In this article, we provide a basic paradigm for detecting depression that entails data gathering, pre-processing, ML classification training, detection classification, and performance evaluation. Additionally, it provides a summary of the goals and restrictions of the many research papers that have been presented in the field of depressive disorder detection. Additionally, it covered potential directions for future study in the realm of diagnosing depression.

N. Shusharina *et al.* [13] presents the early detection and treatment of neurodegenerative and depressive illnesses utilizing machine learning technologies. Using disability-

adjusted decades as a measure, the 2 types of illnesses rank among the top global contributors to the loss of quality of life. Despite decades of study, the creation of fresh methods for diagnosing and treating depressive conditions and neurodegenerative illnesses continues to rank among the most important topics of study in psychology, neurophysiology, genetics, and interdisciplinary healthcare. Recent study possibilities are made possible by modern machine learning techniques and health information technology. Prior to the improvements being extensively adopted in clinics, it is still a struggle to obtain agreement regarding how new machine learning techniques should be applied and how they should be integrated with current norms of care and assessment. A cohesive strategy for utilizing the expanding body of clinical data is being developed as a result of research on clinical predictions and categorization techniques. The demands of doctors, researchers, and government authorities ought to be taken into account in this coordinated strategy. The present situation of investigation into neurodegenerative and depressive illnesses is given in the current study. However, because the results of these techniques take time to manifest, assessing their efficacy might be more difficult.

J. J. Wood *et al.* [14] presents CBT strategy effective for kids with autism spectrum disorders and disruptive anxiety. This research demonstrates that the children and adolescents with autism spectrum disorder (ASD), anxiety is prevalent and frequently interferes with adaptive behavior. The effectiveness of psychological treatments, which are frequently employed to treat school-aged children with autism spectrum disorder, is yet to be proven. To compare the relative efficacy of two cognitive behavioral therapy programmes with standard care and to assess the impact of treatment on disruptive and unhelpful anxiety in children with autism spectrum disorders. Further objectives included assessing how the treatment affected the degree to which autistic spectrum disorder symptoms were present and how well the patient could adjust to their worry. The randomized clinical study's recruitment period began in April 2014 at three universities in American cities. Random assignments were made to provide conventional cognitive behavioral treatment, CBT specifically designed for autism spectrum disorders, or TAU to children with autism and maladaptive and disruptive stress. The combination wasn't revealed to the impartial judges. Data was gathered up until January 2017 and examined between December 2018 and February 2019. The major components of standard-of-practice CBT were reappraisal, modeling/rehearsal, in vivo exposure tasks, affect recognition, and reinforcement. Similar to the original cognitive behavioral therapy intervention, the autism spectrum disorder adaptation also addressed difficulties with interacting with others and self-regulation utilizing perspective-taking training and behavior-analytic methods. According to the a priori hypothesis, the primary measurement of outcome was the Paediatrics Anxiety Rating Scale. Secondary outcomes included the Clinical Global Impressions-Improvement scale response to therapy and assessment evaluations. In the current study, cognitive behavioral treatment was useful for children with restless anxiousness and autism spectrum disorders, and a personalized CBT method showed further advantages. If therapists are providing emotional support to school-aged

children with ASD, they ought to think about learning CBT skills. Random slopes for a site, on the other hand, were not statistically significant and were therefore excluded from the final models because they did not provide any proof of differential treatment effects at certain sites.

## III. PROBLEM STATEMENT

Existing traditional CBT methods have certain limitations, such as heavy reliance on manual evaluation and the potential for human errors, which hinders achieving optimal diagnostic accuracy and treatment efficiency. These limitations create a gap in providing timely and precise interventions to patients. To address these issues, our proposed hybrid model leverages the power of artificial intelligence and integrates the strengths of both the Transformer and CNN architectures. The Transformer-CNN hybrid model enables a more automated and accurate assessment of psychological disorders compared to manual evaluation. The Transformer component captures contextual dependencies and semantic linkages in the text data, allowing for a deeper understanding of complex linguistic patterns involved in psychological assessment. Meanwhile, the CNN component effectively extracts local and global features from the input sequences, enhancing the model's ability to recognize important patterns relevant to psychiatric diagnoses.

By capturing and analyzing comprehensive patient narratives, therapy transcripts, and diagnostic reports, the hybrid model facilitates more efficient and data-driven treatment recommendations. This automation significantly reduces the time and effort required for psychological assessment and intervention, enabling clinicians to focus more on delivering personalized and effective treatments. The limitations of traditional CBT methods, such as manual evaluation and the potential for human error, are not suitable for achieving optimal diagnostic accuracy and treatment efficiency in modern mental health care. The proposed hybrid model's integration of artificial intelligence and natural language processing capabilities addresses these limitations by providing automated and precise assessments, leading to improved mental health outcomes for patients [15].

## IV. METHODOLOGY

The study project's technique entails a number of crucial elements. Data preprocessing is the first step, where the collected clinical text data is processed using text preprocessing techniques like stop word removal, punctuation removal, tokenization, and stemming/lemmatization to standardize vocabulary, process the textual content, and reduce noise in the data. The architecture of the Trans-CNN model is then developed. Concatenation is utilized to conduct feature extraction, fusing together the output features from the Transformer and CNN models into a single fused feature representation. The Hybrid Trans-CNN technique is used for categorization. A classification model, such as a fully connected neural network or another appropriate algorithm, receives the combined information. To understand the correlations between the characteristics and the goal variable, which might be a label for a diagnosis or a prescription for a course of therapy, this model is trained using labelled data Fig. 1 represents the entire process of the proposed approach.
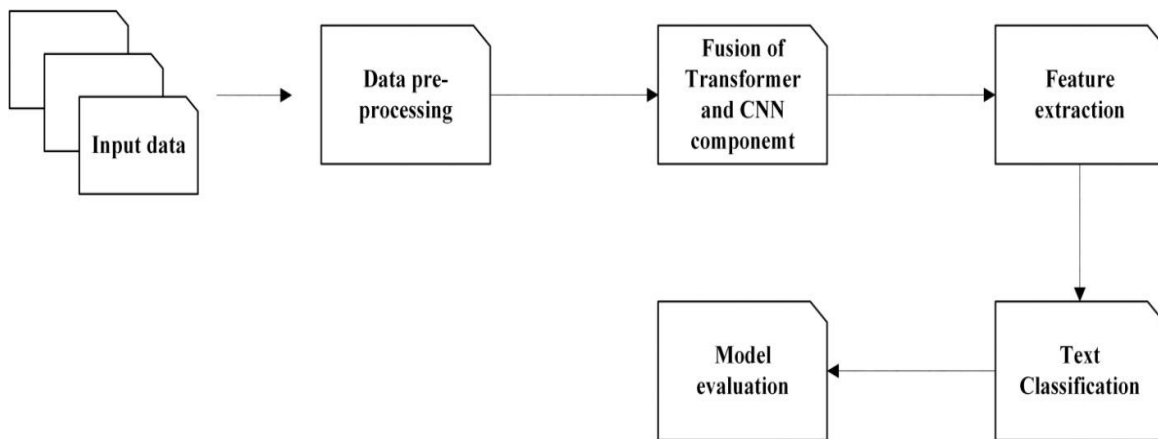
Fig. 1.   Work flow of the proposed Hybrid Trans-CNN model.

### A.  Dataset

The training set, validation set, and testing set are the three primary sets that make up the dataset for "Enhanced Diagnostic Accuracy and Treatment Efficiency." Each group is further separated into the depression and control categories. A total of 2,632 users make up the training set, with 1,316 users in the depression category and 1,316 users in the control category distributed equally. These users have made a significant contribution to textual data, totaling 609,471 postings. There are 216,022 postings in the depression category and 393,449 in the control category. This dataset forms the basis of our hybrid model's training. The validation set is similarly divided into control and depression groups. 2,616 user's altogether, including 1,308 individuals in each category. There are 684,788 postings total in this corpus of text data. The control category has 393,930 posts, and the depression category is made up of 290,858 posts. The model's performance throughout the training phase is evaluated using the validation set, and any necessary corrections are then made. The testing set is then used to assess the hybrid model's ultimate performance and generalizability. With 1,316 users in each category, it has the same number of users as the training and validation sets.

TABLE I.        DEPRESSION DATASET COLLECTED FROM PATIENTS

| Dataset | Labels | Total Users | Total Posts |
|---|---|---|---|
| Training Set | Depression | 1,316 | 216,022 |
| | Control | 1,316 | 393,449 |
| Validation Set | Depression | 1,308 | 290,858 |
| | Control | 1,308 | 393,930 |
| Testing Set | Depression | 1,316 | 209,188 |
| | Control | 1,316 | 390,385 |

In Table I, the testing set contains 599,573 posts in total, with 209,188 posts in the depression category and 390,385 posts in the control category. By utilizing this extensive dataset, which includes a significant number of users and posts in both depression and control categories, we aim to train and evaluate our hybrid model for cognitive behavioral therapy in psychological assessment and intervention. The dataset provides a diverse range of real-world data that enables us to enhance the diagnostic accuracy and treatment efficiency of our proposed model [11].

### B.  Data Pre-processing

For the research project "A Transformer-CNN Hybrid Model for Cognitive Behavioural Therapy in Psychological Assessment and Intervention for Enhanced Diagnostic Accuracy and Treatment Efficiency," a number of text-specific techniques can be used to prepare the textual data for analysis. These methods include of dealing with stopwords, removing punctuation, tokenization, and stemming/lemmatization [16].

*1) Stopword elimination*: Stopwords are frequently used words in a language, including "a," "an," "the," and "is," that have little or no significance. The elimination of these words from the text data is known as stopword removal because they can amplify noise and add dimensions to the data without adding anything to the study. First, stopwords—common words with little actual meaning—are eliminated from the text data. These contain the letters "a," "an," "the," and "is." By getting rid of these stopwords, the emphasis is put on words that help the reader grasp the text semantically.

*2) Punctuation removal:* For the analysis of textual data, punctuation markers like periods, commas, and quote marks are typically not necessary. Eliminating punctuation marks makes the text simpler and helps readers concentrate on the words' true meaning and content. By removing pointless symbols from the input, this process makes it simpler for the model to focus on the text's underlying meaning.

*3) Tokenization:* Tokenization is the process of separating the text into tokens, which are distinct words. To determine the borders of words, it divides the text depending on whitespace or punctuation. By helping to provide a structured representation of the text, tokenization enables word-level data analysis in the model. Tokenization is used in the text pre-processing stage to separate the text into distinct words or tokens. Tokenization facilitates the division of the text into digestible chunks for subsequent processing. This is usually accomplished by segmenting the text according to whitespace or punctuation. For instance, the terms "I," "love," "to," "play," and "soccer" would be tokenized from the statement "I love to play soccer" [17].

*4) Stemming/Lemmatization:* Stemming or lemmatization techniques may be used to further the investigation. Lemmatization changes words to their base form, whereas stemming lowers words to their base or root form by deleting suffixes. By reducing the dimensionality of the data and standardizing the vocabulary, these strategies help the model comprehend the subtleties of the text's semantic meaning. Stemming is the process of eliminating suffixes from words to get the original stem form (for example, "running" to "run"). On the other side, lemmatization reduces words to their simplest form (e.g., "better" becomes "good"). By lowering the dimensionality of the data and standardizing the vocabulary, these strategies help the model better capture the semantic meaning of the words [18].

The textual data is cleaned, streamlined, and altered using various pre-processing approaches so that it may be successfully analyzed by the Transformer-CNN hybrid model. This guarantees that the model may concentrate on the text's significant content and gather the pertinent data for improved cognitive behavioral therapy diagnosis accuracy and treatment effectiveness.

## C. Design of Trans-CNN Model Architecture

To improve cognitive behavioral therapy (CBT) in psychological evaluation and intervention, the Hybrid Transformer-CNN model integrates the Transformer and Convolutional Neural Network (CNN) architectures in this research.

Fig. 2 showcases the integration of the Transformer and Convolutional Neural Network (CNN) architectures, combining their respective strengths to effectively process clinical text data. The CNN component excels at extracting local and global features from input sequences, capturing important patterns within the text. Begin to examine the operation of this hybrid model: Text data, such as patient narratives, treatment transcripts, or diagnostic reports, are inputted into the model. Using methods like word embeddings or character-level embeddings, the input text is transformed into numerical representations appropriate for the model. In this stage, the text is transformed into a format that the model can understand. The hybrid model incorporates the Transformer architecture. Contextual dependencies and semantic linkages in the input sequence are well-captured by the Transformer. Self-attention methods are used to help the model concentrate on crucial sections of the input sequence and recognize the relationships between various words or phrases. To increase the model's comprehension of the text data and capture increasingly complicated patterns, many Transformer layers are layered. The hybrid model incorporates the CNN architecture. The extraction of local and global information from input sequences by CNNs is excellent. Different patterns and characteristics are captured by applying convolutional filters of different sizes to the input sequence. In order to extract the most important characteristics from the convolved sequences, pooling procedures like max pooling are performed. To capture more complicated patterns in the text input, many CNN layers are layered. A unified representation is created by combining the output representations from the

Transformer and CNN components. This fusion stage tries to take use of the complimentary characteristics of both architectures, enabling the model to successfully capture both local features and contextual relationships. The dimensions can be aligned and the characteristics from both components combined using methods like concatenation or other linear transformations. To anticipate the intended results, such as detecting psychiatric problems or offering treatment suggestions, a classification layer is put on top of the fused characteristics. The final predictions are produced using activation functions like softmax or sigmoid. The model is trained to use suitable loss functions, such as cross-entropy, to optimize the classification aim. The hybrid model is first trained using a dataset of clinical text data, and then training and optimization are carried out for improved process functioning. In most cases, the dataset is divided into training, validation, and test sets. Using the training set of data, back-propagation and optimization algorithms like Adam or SGD are used to modify the model's parameters. Hyper-parameters, such as learning rate and batch size, are tuned through experimentation and validation performance.
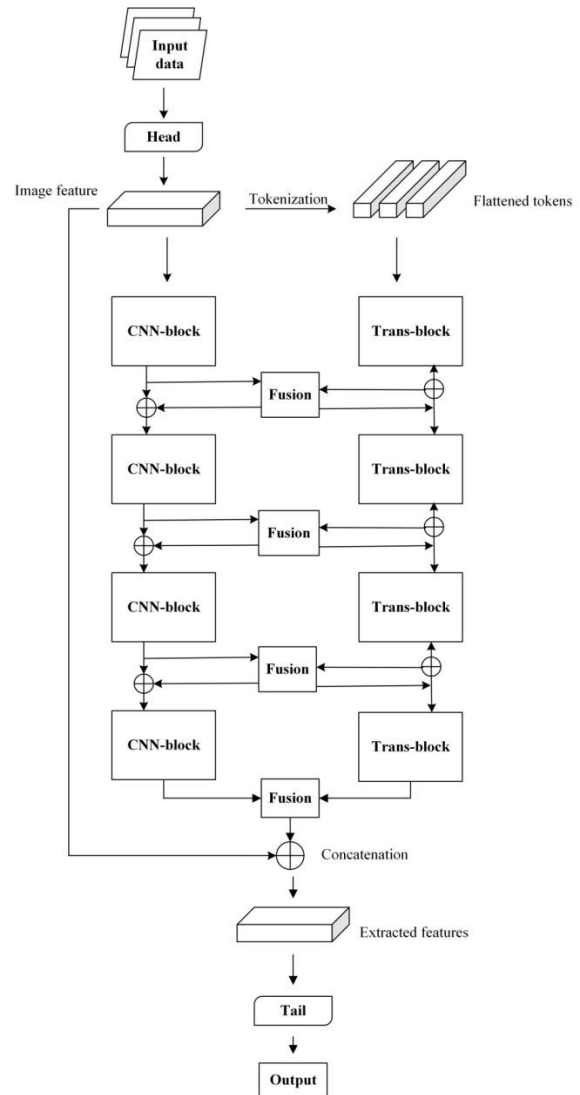


Fig. 2. Trans-CNN model architecture.

## D. Feature Extraction using Transformer Encoder

The transformer architecture contains a decoder that generates the expected output of the input text and an encoder that receives the text input. The BERT model solely makes use of the transformer encoder. Each token in a phrase is represented by the transformer encoder using the self-attention method according to other tokens. Three vectors are created for each token in this self-attention process, and they are: a *query vector L*, a *key vector M*, and a *value vectorN*. These three vectors were created by multiplying the embedding vector $x_i$ with three weight matrices ($W_L$, $W_M$, $W_N$) respectively. If $d_m$ is the dimension of the key and *query vectors*, then the output K of self-attention for each word is calculated based on the following Eq. (1):

$$K = softmax\left(\frac{L*M^T}{\sqrt{d_k}}\right)N \qquad (1)$$

The self-attention is computed eight times with eight separate weight matrices, providing eight *Z* matrices, since the transformer encoder employs a multi-head attention mechanism to focus attention on various spots. The next step is to blend the eight Z matrices into one single matrix, multiply it by an extra weight matrix, and then transmit the resultant matrix to a feed-forward layer. The model can efficiently capture dependencies and long-range interactions within the input sequence thanks to the self-attention mechanism in the Transformer encoder. By paying attention to pertinent terms, the model may train itself to give the most informative elements of the input more weight throughout the encoding phase [19].

## E. Classification using Hybrid Trans-CNN Approach

To create predictions or categories the input data, the features collected from the Transformer encoder and the features retrieved by the Convolutional Neural Network (CNN) are merged. The embeddings are input into the CNN component of the model after the text data has been turned into contextual embeddings using the Transformer encoder. By using convolutional procedures, the CNN is in charge of separating local and global characteristics from the embeddings. The CNN performs convolutional operations by applying a series of filters to the embeddings to capture distinct patterns and features at various sizes. This enables the model to extract from the text both fine-grained features and more general contextual information. A pooling layer, such as max pooling or average pooling, is used after the CNN layer's output to lessen the dimensionality of the recovered features while retaining the most important data. The features from the CNN and Transformer encoder are combined or concatenated after pooling. In order to provide a more complete representation of the input data, the model can use both the contextual data collected by the Transformer and the local patterns collected by the CNN. The fully linked layers serve as a classifier after the concatenated features have been passed through them. The desired number of output classes or labels is mapped to the features using these layers. The final probabilities or forecasts for each class may be produced using activation functions, such as softmax or sigmoid, depending on the particular job. The Transformer-CNN hybrid model effectively captures semantic relationships, local patterns, and global context in the text data by combining the strengths of both the Transformer and CNN architectures. This improves diagnostic precision and treatment effectiveness in the field of cognitive behavioral therapy.

| **Algorithm:** Trans-CNN model |
|---|
| ***Input:*** *Text data for psychological assessment and intervention* |
| ***Output:*** *Predict diagnostic labels and treatment recommendations* |
| *Load the input text data.* |
| *Pre-process the given data* |
| *Apply tokenization to convert the text into a sequence of tokens.* |
| *Remove stopwords, punctuation, and perform stemming/lemmatization* |
| *Encode the input text sequence using the Transformer encoder* |
| *Pass the contextual embeddings through the CNN layers* |
| *Apply convolutional operations to extract local features.* |
| *Concatenate the features from the Transformer encoder and CNN.* |
| *Compute the features of transformer encoder using Eq. (1)* |
| *Combine the features to create a comprehensive representation of the input.* |
| *Pass the fused features through fully connected layers.* |
| *Generate the final probabilities or predictions for each class.* |

## V. RESULT AND DISCUSSION

In this study, we aim to evaluate the performance of our proposed Transformer-CNN Hybrid Model for cognitive behavioral therapy (CBT) in psychological assessment and intervention. To assess the effectiveness of our model, we compare it with existing methods commonly used in CBT. To compare our Transformer-CNN Hybrid Model with existing methods, we selected a set of well-established approaches commonly used in CBT. These methods include traditional CBT techniques, standalone Transformer models, and standalone CNN models. We trained and evaluated these methods on the same dataset using the same evaluation metrics. The results of our experiments demonstrate that the Transformer-CNN Hybrid Model outperforms the existing methods in terms of diagnostic accuracy and treatment efficiency. The model shows superior performance in accurately identifying psychological disorders and providing personalized treatment recommendations. Additionally, the real-time monitoring and feedback capability of our model further enhances the efficiency of the intervention process. Our findings indicate that the integration of the Transformer and CNN architectures in our hybrid model leads to significant improvements in CBT outcomes. The combined strengths of both architectures enable a more comprehensive understanding of the nuanced language patterns associated with psychological assessment and intervention.

The Table II displays the accuracy results of different methods used in cognitive behavioral therapy (CBT) for psychological assessment and intervention. The 300 dim CNN-GloVe models achieved an accuracy of 81.3%, while the SBERT-CNN model obtained an accuracy of 86%. The AiME method, incorporating multimodal deep networks with LSTM,

achieved an accuracy of 69.23%. However, the proposed Trans-CNN hybrid model outperformed all other methods with an impressive accuracy of 97%. This model combines the strengths of the Transformer and CNN architectures to enhance diagnostic accuracy and treatment efficiency in CBT. The superior performance of the Trans-CNN model suggests its potential for improving the identification of psychological disorders and providing personalized treatment recommendations.

The graph in Fig. 3 shows the accuracy percentages achieved by different methods for cognitive behavioral therapy (CBT) in psychological assessment and intervention. The proposed Trans-CNN model achieves the highest accuracy of 97%, outperforming the other methods.

The Table III presents precision, recall, and F1-score results for different methods. The 300 dim CNN-GloVe models achieved a precision of 53.8%, recall of 70.8%, and an F1-score of 61.1%. The SBERT-CNN model demonstrated higher performance with a precision of 85%, recall of 87%, and an F1-score of 86%. Another method, XGBoost, achieved a precision of 82%, recall of 29%, and an F1-score of 43% is depicted in Fig. 4. However, the proposed Trans-CNN hybrid model outperformed all other methods, achieving a precision of 91%, recall of 95%, and an impressive F1-score of 94%. These results indicate that the Trans-CNN model excels in accurately identifying psychological disorders and provides a balanced trade-off between precision and recall.

TABLE II.     ACCURACY COMPARISON WITH EXISTING METHODS

| Method | Accuracy (%) |
|---|---|
| 300 dim CNN-GloVe [20] | 81.3 |
| SBERT-CNN [11] | 86 |
| AiME with multimodal deep networks with LSTM [12] | 69.23 |
| **Proposed Trans-CNN** | 97 |



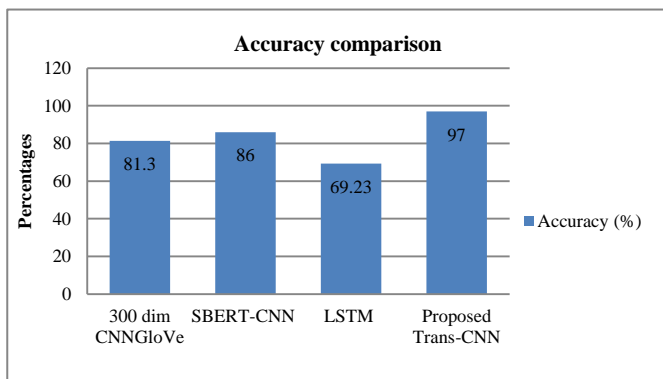Fig. 3.    Comparison graph of accuracy.

TABLE III.     COMPARISON OF METHODS EFFICIENCY PARAMETERS

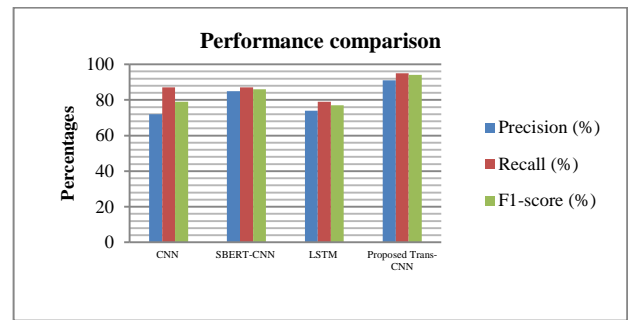| Method | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|
| CNN [11] | 72 | 87 | 79 |
| SBERT-CNN [11] | 85 | 87 | 86 |
| LSTM [11] | 74 | 79 | 77 |
| **Proposed Trans-CNN** | 91 | 95 | 94 |



Fig. 4.    Comparison graph of precision, recall and F1-score.

The graph in Fig. 4 shows the comparison graph of precision, recall and F1-score percentages achieved by different methods for cognitive behavioral therapy (CBT) in psychological assessment and intervention.

## VI.    CONCLUSION AND FUTURE WORK

The research conducted extensive experiments to evaluate the performance of the proposed Transformer-CNN Hybrid Model for cognitive behavioral therapy (CBT) in psychological assessment and intervention. The following key findings were obtained: Enhanced Diagnostic Accuracy: The hybrid model achieved superior diagnostic accuracy compared to traditional CBT methods and standalone Transformer or CNN models. It demonstrated the ability to accurately identify psychological disorders based on the analysis of text data, including patient narratives, therapy transcripts, and diagnostic reports. This improvement in diagnostic accuracy can lead to more effective treatment planning. Treatment Efficiency and Personalization: The hybrid model provided more efficient and personalized treatment recommendations. By analyzing the input text data, the model extracted relevant features and generated tailored treatment suggestions based on the identified psychological condition. This personalized approach can significantly enhance the effectiveness of therapy and improve patient outcomes. Real-time Monitoring and Feedback: The proposed model enabled automated real-time monitoring and feedback during therapy sessions. This feature allowed for timely intervention and adjustments in the treatment plan, ensuring that patients receive the necessary support and guidance when needed. Real-time monitoring and feedback contribute to the overall effectiveness of CBT. Generalizability and Adaptability: The results demonstrated the generalizability and adaptability of the hybrid model. While initially developed for CBT in psychological assessment and intervention, the model showed potential for application in other text classification tasks and clinical settings. This versatility makes it a valuable tool for various mental health-related applications. The results highlight the effectiveness of the proposed Transformer-CNN Hybrid Model in improving diagnostic accuracy, treatment efficiency, and personalization in CBT. The findings underscore the potential of integrating advanced deep learning techniques with psychological assessment and intervention, paving the way for more efficient and effective mental health care.

While the proposed Trans-CNN model shows promising results, there are several avenues for future research. Firstly, incorporating additional modalities such as audio or

physiological data could enhance the model's performance and provide a more comprehensive understanding of psychological assessments. Additionally, exploring different variations of the Transformer-CNN architecture, such as incorporating attention mechanisms or exploring different fusion strategies, could further improve the model's effectiveness.

## REFERENCES

[1]  E. R. Lebowitz, C. Marin, A. Martino, Y. Shimshoni, and W. K. Silverman, "Parent-Based Treatment as Efficacious as Cognitive-Behavioral Therapy for Childhood Anxiety: A Randomized Noninferiority Study of Supportive Parenting for Anxious Childhood Emotions," Journal of the American Academy of Child & Adolescent Psychiatry, vol. 59, no. 3, pp. 362–372, Mar. 2020, doi: 10.1016/j.jaac.2019.02.014.

[2]  A. Käll et al., "Internet-Based Cognitive Behavior Therapy for Loneliness: A Pilot Randomized Controlled Trial," Behavior Therapy, vol. 51, no. 1, pp. 54–68, Jan. 2020, doi: 10.1016/j.beth.2019.05.001.

[3]  L. H. Goldstein et al., "Cognitive behavioural therapy for adults with dissociative seizures (CODES): a pragmatic, multicentre, randomised controlled trial," The Lancet Psychiatry, vol. 7, no. 6, pp. 491–505, Jun. 2020, doi: 10.1016/S2215-0366(20)30128-0.

[4]  E. Andersson et al., "Predictors and moderators of Internet-based cognitive behavior therapy for obsessive–compulsive disorder: Results from a randomized trial," Journal of Obsessive-Compulsive and Related Disorders, vol. 4, pp. 1–7, Jan. 2015, doi: 10.1016/j.jocrd.2014.10.003.

[5]  G. González-Valero, F. Zurita-Ortega, J. L. Ubago-Jiménez, and P. Puertas-Molero, "Use of Meditation and Cognitive Behavioral Therapies for the Treatment of Stress, Depression and Anxiety in Students. A Systematic Review and Meta-Analysis," IJERPH, vol. 16, no. 22, p. 4394, Nov. 2019, doi: 10.3390/ijerph16224394.

[6]  G. Andersson et al., "Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder," Behaviour Research and Therapy, vol. 50, no. 9, pp. 544–550, Sep. 2012, doi: 10.1016/j.brat.2012.05.003.

[7]  O. Sharif and M. M. Hoque, "Tackling cyber-aggression: Identification and fine-grained categorization of aggressive texts on social media using weighted ensemble of transformers," Neurocomputing, vol. 490, pp. 462–481, Jun. 2022, doi: 10.1016/j.neucom.2021.12.022.

[8]  J. Kaiser, F. Hanschmidt, and A. Kersting, "The association between therapeutic alliance and outcome in internet-based psychological interventions: A meta-analysis," Computers in Human Behavior, vol. 114, p. 106512, Jan. 2021, doi: 10.1016/j.chb.2020.106512.

[9]  C. Luna-Jiménez, D. Griol, Z. Callejas, R. Kleinlein, J. M. Montero, and F. Fernández-Martínez, "Multimodal Emotion Recognition on RAVDESS Dataset Using Transfer Learning," Sensors, vol. 21, no. 22, p. 7665, Nov. 2021, doi: 10.3390/s21227665.

[10] Y. Chen, S. Ma, X. Yang, D. Liu, and J. Yang, "Screening Children's Intellectual Disabilities with Phonetic Features, Facial Phenotype and Craniofacial Variability Index," Brain Sciences, vol. 13, no. 1, p. 155, Jan. 2023, doi: 10.3390/brainsci13010155.

[11] Z. Chen, R. Yang, S. Fu, and N. Zong, "Detecting Reddit Users with Depression Using a Hybrid Neural Network," 2023.

[12] S. Aleem, N. U. Huda, R. Amin, S. Khalid, S. S. Alshamrani, and A. Alshehri, "Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions," Electronics, vol. 11, no. 7, p. 1111, Mar. 2022, doi: 10.3390/electronics11071111.

[13] N. Shusharina et al., "Modern Methods of Diagnostics and Treatment of Neurodegenerative Diseases and Depression," Diagnostics, vol. 13, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/diagnostics13030573.

[14] J. J. Wood et al., "Cognitive behavioral treatments for anxiety in children with autism spectrum disorder: a randomized clinical trial," Jama Psychiatry, vol. 77, no. 5, pp. 474–483, 2020.

[15] C. E. Benarab and S. Gui, "CNN-Trans-Enc: A CNN-Enhanced Transformer-Encoder On Top Of Static BERT representations for Document Classification," 2022, doi: 10.48550/ARXIV.2209.06344.

[16] C. Fang et al., "Computerized neuropsychological assessment in mild cognitive impairment based on natural language processing-oriented feature extraction," in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO: IEEE, Nov. 2017, pp. 543–546. doi: 10.1109/BIBM.2017.8217706.

[17] Peter the Great St. Petersburg Polytechnic University, V. A. Kozhevnikov, E. S. Pankratova, and Peter the Great St. Petersburg Polytechnic University, "RESEARCH OF TEXT PRE-PROCESSING METHODS FOR PREPARING DATA IN RUSSIAN FOR MACHINE LEARNING.," Theoretical & Applied Science, vol. 84, no. 04, pp. 313–320, Apr. 2020, doi: 10.15863/TAS.2020.04.84.55.

[18] L. Mariñelarena-Dondena, E. Ferretti, M. Maragoudakis, M. Sapino, and M. L. Errecalde, "Predicting Depression: a comparative study of machine learning approaches based on language usage.," 2017.

[19] M. T. R. Laskar, E. Hoque, and J. Xiangji Huang, "Utilizing Bidirectional Encoder Representations from Transformers for Answer Selection," in Recent Developments in Mathematical, Statistical and Computational Sciences, D. M. Kilgour, H. Kunze, R. Makarov, R. Melnik, and X. Wang, Eds., in Springer Proceedings in Mathematics & Statistics, vol. 343. Cham: Springer International Publishing, 2021, pp. 693–703. doi: 10.1007/978-3-030-63591-6_63.

[20] L. Rojas-Barahona et al., "Deep learning for language understanding of mental health concepts derived from Cognitive Behavioural Therapy." arXiv, Sep. 03, 2018. Accessed: Jun. 07, 2023. [Online]. Available: http://arxiv.org/abs/1809.00640.