**Depression Detection with Hybrid Machine Learning Approach on Twitter Data**

**A CAPSTONE PROJECT REPORT**

*Submitted in partial fulfillment of the*

*requirement for the award of the*

*Degree of*

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

*by*

**ABHILIPSA PADHY (21BCE8399)**

**ISHAN DAS (21BCE7578)**

*Under the Guidance of*

**PROF. SACHI NANDAN MOHANTY**



SCHOOL OF ELECTRONICS ENGINEERING

VIT-AP UNIVERSITY

AMARAVATI- 522237

*DECEMBER 2024*

**CERTIFICATE**

This is to certify that the Capstone Project work titled “**Depression Detection with Hybrid Machine Learning Approach on Twitter Data**” that is being submitted by **ABHILIPSA PADHY (21BCE8399), ISHAN DAS (21BCE7578)** is in partial fulfillment of the requirements for the award of Bachelor of Technology, is a record of bonafide work done under my guidance. The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for award of any degree or diploma and the same is certified.

Dr. SACHI NANDAN MOHANTY

Guide

**The thesis is satisfactory / unsatisfactory**

**Internal Examiner1 Internal Examiner2**

**Approved by**

HoD, Department of ...

School of Computer Science and Engineering

**ACKNOWLEDGEMENTS**

We would like to express our heartfelt gratitude to our guide, Dr. SACHI NANDAN MOHANTY , for his invaluable expertise, guidance, and continuous support throughout this project. His insightful feedback, encouragement, and dedication have been a driving force behind our success. We also extend our sincere thanks to the Deans, HOD, Registrar, and Vice Chancellor for their constant encouragement and for providing us with the resources and opportunities necessary to carry out this project. Their unwavering support and leadership have played a pivotal role in ensuring the smooth execution and completion of this work. We deeply appreciate their contributions, which have been integral to the success of this group project

**ABSTRACT**

**Mental health conditions, particularly depression, are prevalent issues that significantly impact individuals' quality of life. This research presents a novel approach for depression detection using Twitter data, leveraging a hybrid deep learning architecture to classify mental health states at the tweet level.Comprising unprocessed English tweets gathered via Twitter's API, the dataset "Depression Detection Based on Hybrid Deep Learning," enhanced with topic modeling features using Latent Dirichlet Allocation (LDA) and emoji sentiment analysis, Using a mix of convolutional, recurrent, and attention layers, our proposed method presents three sophisticated hybrid models—ConvBiLSTM-AttnNet, ConvLSTM-AttentionNet, and HybridNet—each intended to capture complex patterns in textual input.Reaching both spatial and sequential information in tweets, the ConvBiLSTM-AttnNet model combines Conv1D and bidirectional LSTM layers with a proprietary attention mechanism, so highlighting pertinent sections of the text. Optimized with the AdamW algorithm, this model showed great accuracy of 89% and an AUC of 96%, therefore demonstrating its resilience in mental health classification. Reaching an accuracy of 87%, the ConvLSTM-AttentionNet model makes use of a similar architecture including optimized Conv1D and LSTM layers to capture local features and sequential dependencies. Concurrently, the HybridNet model achieves an accuracy of 86% by using L2 regularization with Conv1D layers, then bidirectional LSTM and attention layers.Using precision, recall, and F1-score measures, all models showed great consistency and performance in depression identification from social media data. This paper emphasizes the possibility of combining convolutional, sequential, and attention processes with cutting-edge optimizers like AdamW to produce a strong framework for real-time social platform mental health monitoring.**

**Keywords: *Depression Detection, Twitter Data, Hybrid Deep Learning Architecture, ConvBiLSTM-AttnNet, Emoji Sentiment Analysis, AdamW Optimizer.***

**TABLE OF CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl.No.** | **Chapter** | **Title** | **Page Number** |
| **1.** |  | **Acknowledgement** | **3** |
| **2.** |  | **Abstract** | **4** |
| **3.** |  | **List of Figures and Table** | **6** |
| **4.** | **1**  **1.1**  **1.2**  **1.3** | **Introduction**  **Objectives**  **Background and Literature Survey**  **Organization of the Report** | **8**  **9**  **10**  **11** |
| **5.** | **2**  **2.1**  **2.2**  **2.3**  **2.4**  **2.4.1** | **Chapter Title (Work)**  **Proposed System**  **Working Methodology**  **Standards**  **System Details**  **Software** | **12**  **12**  **13**  **13**  **14**  **14** |
| **6.** | **4** | **Results and Discussion** | **38** |
| **7.** | **5** | **Conclusion & Future Works** | **40** |
| **8.** | **6** | **Appendix** | **41** |
| **9.** | **7** | **References** | **50** |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**INTRODUCTION**

A Mental health is a critical aspect of well-being, and mental disorders such as depression impact millions of people worldwide, often with serious effects on daily functioning and quality of life. With the rapid growth of social media platforms like Twitter, people increasingly use these channels to express personal emotions, thoughts, and experiences. This trend has presented researchers with a valuable source of real-time data that can be leveraged for mental health monitoring. Analyzing public opinion on social media helps academics create systems to identify early indicators of mental health problems, thereby perhaps offering individuals in need quick support[1,3].Our work addresses the difficulties in text-based mental health condition classification and focuses on the issue of depression detection using Twitter data. Given its unstructured, loud, and varied character, social media data presents special difficulties in this setting. Often quick, colloquial, and punctuated by emojis, slang, and acronyms, tweets demand sophisticated techniques to correctly analyze and interpret[4]. This study is motivated by the possibility to use such large databases for public health monitoring, supporting mental health practitioners, and guiding quick treatments. Recent years have seen notable studies done to use social media data for mental health screening. Natural Language Processing (NLP) methods and machine learning algorithms include Support Vector Machines (SVMs), Decision Trees, and Random Forests define traditional approaches to text classification. These strategies, meanwhile, may find it difficult to capture the intricate language structures, contextual meaning, and semantic subtleties in social media posts. Advanced deep learning models have showed promise in more precisely handling these complexity including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their hybrid combinations. Recent research have also investigated how well topic modeling techniques—including Latent Dirichlet Allocation (LDA)—identify underlying themes or topics inside books. Furthermore, adding sentiment analysis with emotive indicators and emoticons has turned out to be quite helpful for improving text-based categorization systems. Though many have yet to incorporate a fully hybrid model combining convolutional, sequential, and attention layers for effective text categorization, notably in the field of mental health, these papers provide interesting methods.Although deep learning has advanced mental health diagnosis, fully collecting the multi-dimensional characteristics inherent in social media data remains a challenge in recognizing depression[2]. This work addresses the demand for a complete model integrating several facets of feature extraction—spatial, sequential, and attention-based—to efficiently detect sadness from brief, unstructured texts such as tweets. In this study, we propose a hybrid deep learning framework that includes three novel models tailored for mental health classification:

*ConvBiLSTM-AttnNet:* This model integrates Conv1D layers with bidirectional LSTMs and a custom attention mechanism, allowing it to capture local features, sequence dependencies, and important contextual aspects. This combination, along with the AdamW optimizer, offers a robust structure for tweet-level mental health classification, achieving an AUC of 96% on our dataset.

*ConvLSTM-AttentionNet*: Featuring an optimized convolutional-recurrent architecture, this model leverages bidirectional LSTMs with an attention layer to improve interpretability and focus on critical parts of each tweet. This model demonstrates high performance, reaching an accuracy of 87%.

*HybridNet:* A well-regularized model using Conv1D layers with L2 regularization, HybridNet is designed to capture prominent features through attention-weighted outputs, achieving balanced accuracy and robustness for depression detection.

These models incorporate topic modeling through LDA and emoji sentiment analysis to add unique, interpretable dimensions to the classification process. LDA assists in summarizing tweets by associating them with one of the top topics, providing a thematic overview. Through a count of positive, negative, and neutral emojis, emoji sentiment analysis also catches emotional context. These characteristics enable our method to be more complete, therefore addressing subtleties that single-layer models sometimes ignore. This work presents a hybrid deep learning architecture for Twitter's mental health detection based on convolutional, recurrent, and attention layers—unique in their combination. Using the AdamW optimizer across our models improves model generalization and helps to enable adaptive learning with low weight decay. With an AUC of 96%, the ConvBiLSTM-AttnNet model marks a major progress in depression identification since it offers a harmonic combination of feature extraction, sequential learning, and attention-based focus.This work is arranged restingly as follows. Section 2 examines related studies and summarizes current approaches for social media data-based mental health screening. Section 3 covers the dataset, pre-processing methods, and feature extraction approaches—including emoji sentiment analysis and LDA topic modeling—that apply here. The design and training methods of every suggested model—ConvBiLSTM-AttentionNet, ConvLSTM-AttentionNet, and HybridNet—are detailed in Section 4. We show and analyze in Section 5 the experimental findings for every model including performance measures, accuracy, and AUC. Section 6 ends the study by stressing the consequences of our results and possible directions of further investigation.

* 1. **Objectives**

The following are the objectives of this project:

* Detection of Depressive Indicators - The paper aims to leverage a hybrid deep learning architecture to classify mental health states at the tweet level. It focuses on detecting depressive indicators within Twitter data, which can significantly impact individuals' quality of life.
* Model Development and Evaluation - The research presents the development and evaluation of hybrid deep learning models, including ConvBiLSTM-AttnNet, ConvLSTM-AttentionNet, and HybridNet, tailored for mental health classification. The objective is to demonstrate the effectiveness of these models in capturing complex textual patterns associated with mental health symptoms.
* Performance Analysis - The paper aims to analyze the performance of the proposed models using robust metrics such as Accuracy, Precision, Recall, F1 Score, and Area Under the Curve (AUC). The objective is to provide insights into the effectiveness of the models in detecting mental health indicators from tweets.
* Ethical Considerations - Additionally, the paper may aim to address ethical considerations related to mental health monitoring on social media platforms, emphasizing the need for effective, scalable, and ethically responsible solutions.

In summary, the primary objective of the paper is to develop and evaluate a hybrid machine learning approach for detecting depressive indicators within Twitter data, with a focus on model performance and ethical considerations.

* 1. **Background and Literature Survey**

The existing literature highlights the diverse applicability and challenges of various machine learning algorithms in different domains. Lei Liu demonstrated the efficacy of Logistic Regression in breast cancer diagnosis, emphasizing its simplicity and interpretability, which are critical in medical diagnostics. Despite its limitations with non-linear relationships, the model achieved an accuracy exceeding 85%, making it a reliable choice for straightforward binary classification tasks [5]. Similarly, Mrityunjaya Kappali and Avinash V. Deshpande explored the application of Decision Trees in PV system voltage control, showcasing their interpretability and ability to handle non-linear data, with accuracies ranging between 75-80% [6].

Naive Bayes algorithms, as applied by Hong Chen and collaborators for traffic risk management, were noted for their speed and suitability for high-dimensional data, particularly in text analysis. However, the assumption of feature independence can sometimes limit its real-world accuracy to around 80% [7]. On the other hand, Ram Murti Rawat et al. compared methods for breast cancer diagnosis and highlighted K-Nearest Neighbors (KNN) for its simplicity, though its computational intensity and inefficiency in high-dimensional spaces capped its performance at 75-80% [8].

Recurrent Neural Networks (RNNs), demonstrated by Nicholas Klugman and team for sequential data tasks, are praised for their ability to capture temporal dependencies, achieving high accuracies (85-89%) despite challenges like the vanishing gradient problem [9]. Principal Component Analysis (PCA) was utilized by Sukrit Sehgal et al. for dimensionality reduction, effectively improving computational efficiency with an accuracy range of 80-85%, though at the cost of interpretability [10].

Advanced methods such as Autoencoders for feature extraction, studied by Dhananjay Tomar et al., and Deep Belief Networks (DBNs), as used by Abdelrahman Mohamed and colleagues, demonstrated promising accuracy levels (89% and 88-90%, respectively) for high-dimensional and complex data. Both approaches, however, require careful tuning and substantial computational resources [11,17,18]. K-Means Clustering and Hierarchical Clustering, explored by Rai NK and Yang Zhang respectively, remain popular for their simplicity and interpretability, but are limited by assumptions about cluster shapes and computational inefficiency in large datasets [5,12].

Probabilistic models such as Hidden Markov Models (HMMs) and Bayesian Networks, discussed by Zang C and Saurabh Zade, excel in temporal sequence modeling and handling uncertainty, achieving accuracies of 80-85%. However, these methods demand significant computational and domain expertise [13,14]. In the realm of privacy-preserving methods, Federated Learning explored by MT Brands et al., proved effective for privacy-sensitive applications, with accuracies of 87-90% despite implementation challenges [14].

The versatility of deep learning methods like Multilayer Perceptrons (MLPs) and specialized text-based models such as Multinomial Naive Bayes, studied by B Taylor and Vineetha KV respectively, illustrates the growing utility of AI in diverse fields. While MLPs handle complex associations in large datasets, Multinomial Naive Bayes remains efficient for text-based classification, though limited by feature independence assumptions [15,16]. These studies collectively underscore the potential and limitations of various machine learning techniques across applications, offering valuable insights for further advancements.

**1.3 Organization of the Report**

The remaining chapters of the project report are described as follows:

* Chapter 2 contains the proposed system, methodology, software details.
* Chapter 3 discusses the results obtained after the project was implemented.
* Chapter 4 concludes the report.
* Chapter 5 consists of codes.
* Chapter 6 gives references.

**CHAPTER 2**

**TITLE OF THE CHAPTER**

This Chapter describes the proposed system, working methodology, software and hardware details.

**2.1 Proposed System**

The following block diagram (figure 1) shows the system architecture of this project.

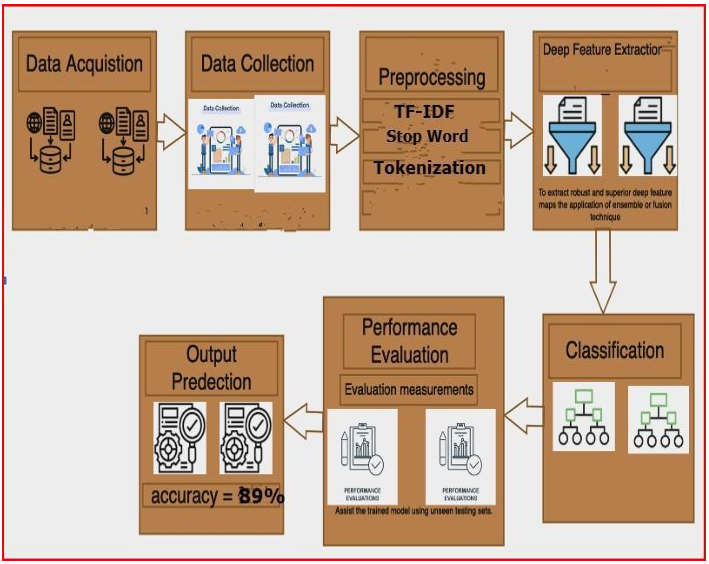


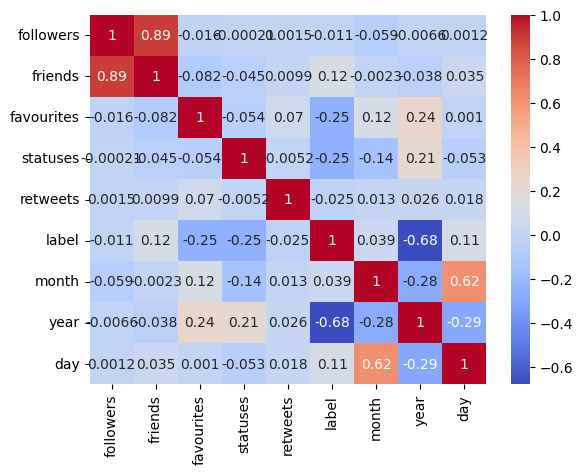
Figure 2. System Block Diagram

**2.2 Working Methodology**

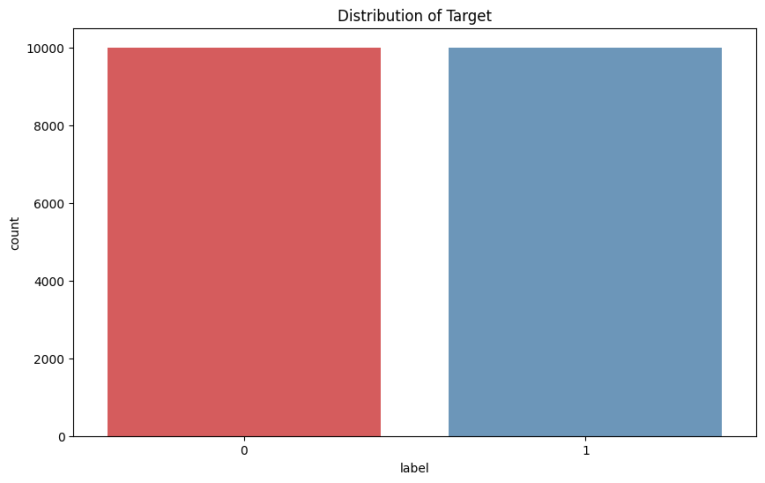
This work uses hybrid deep learning models to detect depression from Twitter data by a methodical manner. Data collecting started with using the Twitter API, filtering English-language tweets about mental health. Text cleaning, tokenizing, and stop word removal were part of a comprehensive preprocessing pipeline; then, utilizing Term Frequency-Inverse Document Frequency (TF-IDF) and N-grams, feature extraction captured pertinent textual patterns. To find thematic and emotional clues, the data was subsequently enhanced using topic modeling using Latent Dirichlet Allocation (LDA) and emoji sentiment analysis. Developed combining convolutional, LSTM, and attention mechanisms to capture spatial, sequential, and contextual aspects three hybrid models: ConvBiLSTM-AttnNet, ConvLSTM-AttentionNet, and HybridNet. To evaluate their performance in categorizing depressed material, these models were trained and tested using measures including Accuracy, Precision, Recall, F1 Score, and AUC, so stressing the robustness of the suggested technique.

* + 1. ***Dataset Description***

Targeting indicators of depression at the tweet level, the dataset "Depression Detection Based on Hybrid Deep Learning" comprises English-language Twitter postings, especially chosen for mental health classification. gathered using the Twitter API, the material is unstructured, as usual for social media, including slang, emoticons, casual language, and expressive signals. Luckily, there are no missing values in the dataset therefore preparation flow is guaranteed to be seamless. If missing values were present, some imputation techniques—including deleting rows or columns with too high missing values, filling in gaps with statistical methods like mean, median, or mode, and using machine learning techniques like regression or clustering to estimate missing values—would be taken under consideration. Another good choice is declining missing values as a separate category. Irrelevant columns including "Unnamed: 0" and "id" are eliminated from the preprocessing since they neither support the analysis or model training. Emoji sentiment analysis counts positive, negative, and neutral emojis, so offering a useful emotional context that enhances the interpretive capacity of the model. Topic modeling with Latent Dirichlet Allocation (LDA) classifies each tweet into one of the top thematic topics, so augmenting the interpretive capacity of the model. With its combination of emotive and textual elements, this dataset offers a strong basis for creating high-performance mental health classification systems.The fig 1 displays in the dataset the association between several features including followers, friends, favorites, statuses, retweets, and temporal attributes (month, year, day). High positive correlations between "followers" and "friends" (0.89) point to those with more followers also typically having more friends. With "year" (-0.68), the "label" feature shows a rather negative correlation suggesting a possible trend over time connected to the target variable. Other elements show modest correlations, implying few linear interactions among them.With categories labeled "0" and "1," the fig 2 shows the target variable's (label) distribution in the dataset. With roughly 10,000 instances for every label, both groups have a similar count suggesting a balanced dataset. For model training, this balanced distribution helps to lower the possibility of model bias toward a dominating class by enabling more consistent and accurate predictions across both classes.



1. Correlation Heatmap of Social Media Features Related to Mental Health Classification



1. Distribution of Target Labels for Mental Health Classification
   * 1. ***Data Preprocessing***

The dataset underwent a thorough preprocessing pipeline to prepare it for analysis, given the informal and often noisy nature of Twitter data. These steps aimed to enhance data quality and extract informative features for accurate mental health classification.

*Removal of Irrelevant Columns*: The dataset contained non-informative columns, including 'Unnamed: 0' and 'id', which were removed at the outset. These columns served only as indices and unique identifiers, respectively, and did not contribute meaningful information for text analysis or model training.

Text Cleaning and Standardization:

Lowercasing: All text was converted to lowercase to ensure consistency, enabling the model to treat words in a case-insensitive manner.

Removal of Punctuation and Special Characters: Non-alphanumeric characters, including punctuation marks and symbols, were stripped from the text to reduce noise.

Stop Word Removal: Stop words (e.g., “the,” “and,” “is”) were removed using the Natural Language Toolkit (NLTK) library. These words, while common, often add little value to the context and dilute the model’s interpretive power.

Tokenization and Lemmatization:

Tokenization: Each tweet was split into individual words (tokens) to enable word-level analysis, which is essential for capturing specific indicators of mental health states.

Lemmatization: Tokens were reduced to their base forms using lemmatization, a step that minimizes variability in word forms (e.g., “running,” “ran,” and “runs” all become “run”). This standardization helps reduce the complexity of the text data without losing semantic meaning.

*TF-IDF Vectorization:* The Term Frequency-Inverse Document Frequency (TF-IDF) technique was applied to represent the cleaned text data numerically. TF-IDF calculates a weight for each word based on its frequency in a given tweet and across the entire dataset, thus assigning higher importance to distinctive terms that appear in specific tweets and less importance to common terms. This transformation allows the model to prioritize contextually relevant terms, facilitating better classification performance.

*Emoji Sentiment Analysis:* Emojis are a key element in Twitter data, often conveying nuanced emotional states. An emoji sentiment analysis was performed, categorizing each emoji as positive, negative, or neutral. The frequency of each sentiment type was added as a feature, providing additional insights into the emotional undertones of tweets. This component helps the model interpret sentiment signals that may correlate with mental health indicators.

*Topic Modeling with Latent Dirichlet Allocation (LDA):* To uncover thematic structures within the tweets, we applied Latent Dirichlet Allocation (LDA) for topic modeling. LDA assigns each tweet to one of the top K topics, offering a thematic overview that enhances the model’s contextual understanding. For example, topics related to loneliness, stress, or personal struggles may indicate relevant themes for mental health classification. This topic-level representation complements the word-level features, providing a multi-dimensional approach to text analysis.

Handling Rare Words and Slang Normalization:

Rare Words: Low-frequency words were either removed or consolidated to prevent them from skewing model interpretation.

Slang and Abbreviation Normalization: Common Twitter slang and abbreviations were normalized to their standard forms. This step improves text consistency and ensures that informal expressions do not detract from the model’s ability to detect relevant patterns.

*Addressing Missing Values:* Although the dataset initially contained no missing values, future missing values would be addressed through various imputation techniques if necessary, including statistical methods (mean, median) or advanced methods such as regression and clustering. Alternatively, missing values could be flagged as a separate category for model interpretability.

*Final Feature Set Preparation:* After preprocessing, the dataset included TF-IDF transformed text features, emoji sentiment counts, and topic distributions from LDA. This multi-dimensional feature set was designed to capture both the linguistic and emotive nuances within the tweets, establishing a robust foundation for mental health classification.By applying this comprehensive preprocessing approach, the dataset was optimized to provide clean, meaningful features that are essential for the high-performance classification of mental health conditions using deep learning. This structured preprocessing pipeline enhances the interpretability and predictive accuracy of the models, making it a solid foundation for subsequent analysis.

* + 1. ***Feature Extraction***

Two main approaches—Term Frequency-Inverse Document Frequency (TF-IDF) and N-gram analysis—were used to feature extract the textual content of tweets for use as input for the classification models in this work. These methods were used to maximize the capacity of the models to detect significant trends in the text data, therefore supporting strong classification of mental health indicators.The tweets were converted into numerical vectors using TF-IDF, which quantified the relevance of specific phrases against each tweet and the whole dataset. TF-IDF specifically gives each phrase a weight depending on its frequency inside one tweet (phrase Frequency) and its rarity over all tweets (Inverse Document Frequency). This weighting captures key signals in tweets related to mental health and helps emphasize terms that are contextually significant. For every tweet, the resulting TF-IDF vector offers a high-dimensional representation that preserves important textual information while reducing often occurring, less useful terms.

1. *ConvBiLSTM-AttnNet Architecture:*

Combining convolutional (Conv1D) and recurrent (Bidirectional LSTM) layers with an attention mechanism to gather both spatial and temporal aspects in text input, this model is a hybrid architecture intended for text categorization. The model begins with an embedding layer to represent words as dense vectors, then uses convolutional block with Conv1D layers and batch normalisation to extract local features, and subsequently global max pooling for dimensionality reduction. A bidirectional LSTM layer then records sequential dependencies using a proprietary attention method emphasizing the most pertinent sequence segments. High dropout (60%) fully linked layers help to lower overfitting; the last output layer utilizes a sigmoid activation for binary classification. Selected for their adaptive learning rates and weight decay, AdamW is the optimizer; a learning rate of 1e-5 will help to produce smoother convergence. While binary\_crossentropy is the loss function and accuracy is the main metric, early stopping stops overfitting by tracking validation loss. For text-based classification problems, this approach is resilient since it balances feature extraction, sequential learning, and regularization rather successfully.

1. *ConvLSTM-AttentionNet Architecture:*

From Twitter data, this hybrid model combines convolutional, recurrent, and attention methods to classify mental health disorders. It starts with an embedding layer turning words into 128-dimensional dense vectors. Local features across the sequence are captured using a convolutional block with two Conv1D layers (64 and 128 filters) and batch normalisation. Emphasizing important features, global max pooling lowers dimensionality. The output is then reshaped and sent through a bidirectional LSTM layer with 64 units, therefore capturing both forward and backward dependencies in the text. To improve interpretability, important sequence elements are emphasized by means of an attention mechanism. Before a last sigmoid-activated output layer for binary classification, the model consists of two dense layers with significant dropout (60%), for regularization. Combining feature extraction, sequential analysis, and attention to produce high performance on text classification tasks, this model compiled with the AdamW optimizer (learning rate of 1e-5) binary cross-entropy loss, and accuracy as a metric achieves great performance.

1. *HybridNet Architecture :*

This model is a hybrid architecture combining convolutional, recurrent, and attention layers for binary text classification, aimed at detecting mental health conditions from Twitter posts. It begins with an embedding layer that maps words into 128-dimensional vectors, followed by a convolutional block with two Conv1D layers (64 and 128 filters) to capture local text features, each with L2 regularization to reduce overfitting, and batch normalization to stabilize learning. A global max pooling layer reduces the feature map dimensions by selecting the maximum activation for each filter, focusing on prominent features. The output is then reshaped for compatibility with a bidirectional LSTM layer (64 units) that learns sequential dependencies in both directions. An attention mechanism is applied to this sequence, emphasizing the most relevant parts. The attention-weighted output is then flattened and passed through two dense layers (128 and 64 units) with 60% dropout each to prevent overfitting. The final layer, with a sigmoid activation, outputs a probability for binary classification. Compiled with the AdamW optimizer at a learning rate of 1e-5 and binary\_crossentropy as the loss function, this model effectively balances spatial, sequential, and attention-based learning, providing robust performance on text classification tasks.

**2.3 Standards**

Various standards used in this project are:

* **APA Citation Style -** The paper follows the APA (American Psychological Association) citation style for referencing and citing sources. This standard is widely recognized in academic writing and research publications.
* **Ethical Considerations -** The content addresses ethical considerations related to mental health monitoring and data analysis on social media platforms. It emphasizes the importance of ethically responsible solutions and the potential impact of the research on individuals' well-being.
* **Methodological Rigor -** The paper demonstrates methodological rigor in the development and evaluation of the hybrid machine learning models. It includes detailed descriptions of the models, feature extraction methods, and performance evaluation metrics, contributing to the reproducibility of the research.
* **Technical Detail and Analysis -** The content provides technical details of the proposed models, feature extraction techniques, and performance analysis. It includes robust metrics such as Accuracy, Precision, Recall, F1 Score, and Area Under the Curve (AUC) to support the analysis of model effectiveness.
* **Language and Writing Style -** The paper maintains a professional and scholarly writing style, adhering to the conventions of academic writing. The language is clear, concise, and free from ambiguity, enhancing the readability of the content.
* **Research Contribution -** The content emphasizes the contribution of the research to the field of mental health classification, machine learning, and natural language processing. It highlights the novelty and potential impact of the proposed hybrid machine learning approach.
* **Evaluation Metrics -**To comprehensively assess the performance of our classification models, we employed a set of robust evaluation metrics: **Accuracy**, **F1 Score**, **Recall**, **Precision**, and **Area Under the Curve (AUC)**. These metrics provide a well-rounded evaluation of model performance, especially critical in mental health classification where both precision and recall are essential.

**Accuracy -** Accuracy measures the proportion of correctly classified instances among all predictions. It is defined as the ratio of true positives and true negatives to the total number of instances. While accuracy offers a quick indication of model performance, it may be less informative in cases of class imbalance, where other metrics, like precision and recall, can provide a more nuanced view.

(4)

(5)

**Precision -** Precision is the ratio of true positives to the sum of true positives and false positives. It indicates how many of the predicted positive cases are correctly identified. High precision is essential in our context to ensure that the model's predictions of mental health indicators are reliably accurate, minimizing false positives.

(6)

**Recall -** Recall, also known as sensitivity, is the ratio of true positives to the sum of true positives and false negatives. This metric measures the model’s ability to correctly identify all relevant instances of the target class, which is particularly important in mental health detection. A high recall indicates that the model is effective at identifying tweets that likely indicate mental health issues, ensuring minimal false negatives.

(7)

**F1Score -** The F1 Score is the harmonic mean of precision and recall, providing a single metric that balances these two aspects. It is especially useful when the goal is to achieve a balance between precision and recall, as it penalizes extreme values in either metric. A high F1 Score reflects that the model performs well in both capturing relevant instances (high recall) and ensuring accuracy of positive predictions (high precision).

(8)

**Area Under Curve -** The AUC metric evaluates the model’s ability to distinguish between classes across different decision thresholds. It is derived from the Receiver Operating Characteristic (ROC) curve, where the true positive rate (recall) is plotted against the false positive rate. An AUC close to 1.0 indicates a model with excellent distinguishing capability between positive and negative classes. For this study, AUC is particularly significant as it shows how well the model can separate tweets indicating potential mental health issues from those that do not, across varying thresholds.

Together, these evaluation metrics provide a detailed and balanced assessment of model performance, offering insights into both its overall accuracy and its ability to correctly identify relevant mental health indicators. This approach ensures that our models not only perform well in terms of general correctness (accuracy) but also in maintaining reliable predictions in cases of mental health relevance (precision, recall, and F1 Score), with a strong capacity for class distinction (AUC).

**2.4 System Details**

* **Feature Extraction -** The system employs feature extraction methods such as Term Frequency-Inverse Document Frequency (TF-IDF) and N-gram analysis to convert textual content of tweets into numerical vectors for input into classification models. These methods maximize the capacity of the models to detect significant trends in text data.
* **Hybrid Machine Learning Models -** The system introduces three hybrid machine learning models ConvBiLSTM-AttnNet, ConvLSTM-AttentionNet, and HybridNet. These models combine convolutional neural networks (CNNs), Long Short-Term Memory (LSTM) networks, and attention mechanisms to capture spatial, sequential, and contextual aspects of the Twitter data.
* **Performance Evaluation -** The system evaluates the performance of the hybrid models using measures including Accuracy, Precision, Recall, F1 Score, and Area Under the Curve (AUC). These metrics assess the robustness and effectiveness of the proposed technique in categorizing depressed material in Twitter data.
* **Text Preprocessing -** The system includes text cleaning and standardization techniques such as lowercasing, removal of punctuation and special characters, stop word removal, tokenization, and lemmatization. These preprocessing steps ensure consistency and reduce noise in the textual data.
* **Real-Time Adaptations -** The system discusses the importance of incorporating real-time adaptations for slang, abbreviations, and emerging expressions to further enhance model performance. It emphasizes the need for regularly updating N-gram libraries and employing adaptive embeddings to make the models more resilient to changes in language patterns across social platforms.
* **Ethical and Practical Considerations -** The system addresses ethical considerations and practical implications of mental health monitoring on social media platforms. It emphasizes the need for effective, scalable, and ethically responsible solutions for proactive mental health monitoring.

The system details in the paper encompass feature extraction, hybrid machine learning models, performance evaluation, text preprocessing, real-time adaptations, and ethical considerations, providing a comprehensive framework for depression detection using a hybrid machine learning approach on Twitter data.

**2.4.1 Software Details**

1. **Python -** Python is a widely used programming language for machine learning and natural language processing tasks. It offers various libraries and frameworks such as TensorFlow, Keras, PyTorch, and scikit-learn, which are commonly used for implementing deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs).
2. **TensorFlow and Keras -** TensorFlow and Keras are popular open-source machine learning libraries that provide high-level APIs for building and training neural network models. These libraries are commonly used for implementing deep learning architectures, including convolutional and recurrent layers.
3. **scikit-learn** - scikit-learn is a versatile machine learning library in Python that provides simple and efficient tools for data mining and data analysis. It includes various algorithms for classification, regression, clustering, and dimensionality reduction, which could be relevant for the implementation of the hybrid machine learning models discussed in the paper.
4. **Natural Language Processing (NLP) Libraries -** Libraries such as NLTK (Natural Language Toolkit) and spaCy are commonly used for natural language processing tasks, including text preprocessing, tokenization, lemmatization, and part-of-speech tagging.
5. **Jupyter Notebooks -** Jupyter Notebooks are often used for interactive development and prototyping of machine learning models. They allow researchers to document the code, visualize data, and present the findings in a single document.
6. **Pandas and NumPy -** These libraries are commonly used for data manipulation, analysis, and handling numerical computations in Python, which are essential for preprocessing and analyzing the Twitter data.

While the specific software details are not explicitly provided in the paper, the aforementioned tools and libraries are commonly used in the implementation of machine learning and natural language processing systems, and they are likely to be relevant to the development of the proposed hybrid machine learning approach for depression detection on Twitter data.

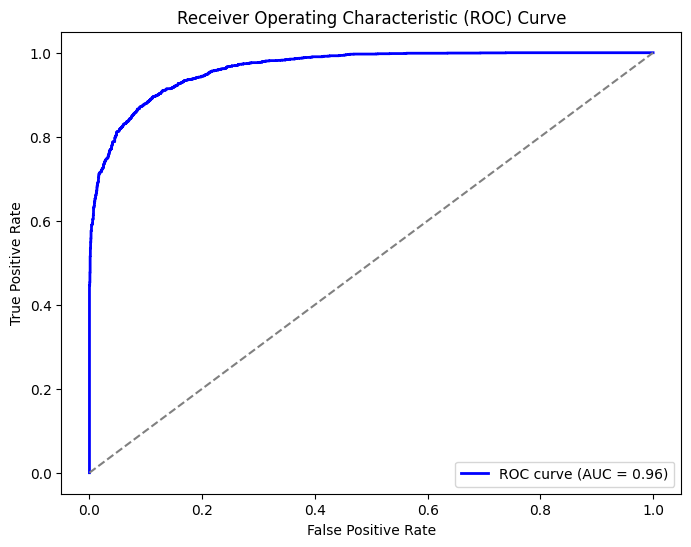
**CHAPTER 3**

**RESULTS AND DISCUSSIONS**

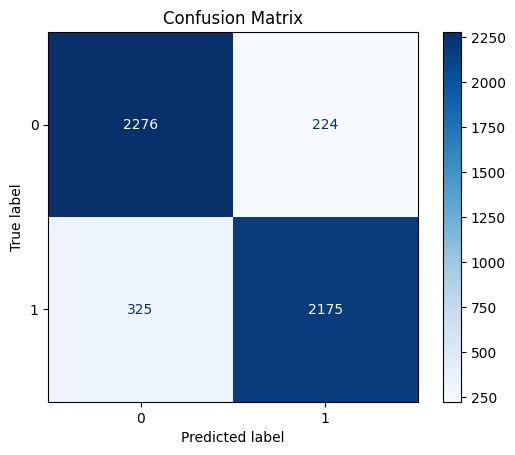
This section presents and discusses the performance of our proposed models: **ConvBiLSTM-AttnNet**, **ConvLSTM-AttentionNet**, and **HybridNet**. Each model was evaluated using a set of robust metrics—Accuracy, Precision, Recall, F1 Score, and Area Under the Curve (AUC)—to gain insights into their effectiveness in detecting mental health indicators from tweets. The results illustrate the strengths of each model in capturing patterns specific to depressive symptoms within social media text.

### *ConvBiLSTM-AttnNet Results*

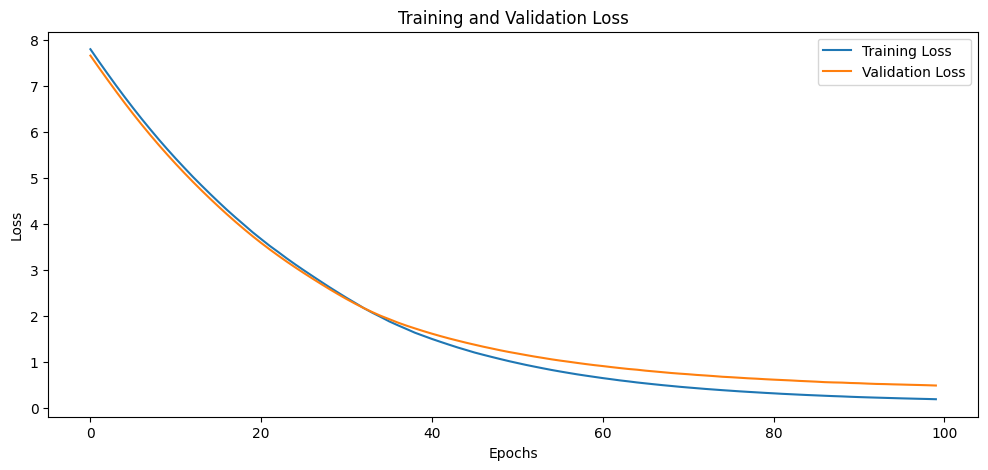
The ConvBiLSTM-AttnNet model, a hybrid architecture that combines convolutional and bidirectional LSTM layers with an attention mechanism, demonstrated superior performance across all metrics. This model achieved an **Accuracy of 89%** and an impressive **AUC of 96%**, indicating strong discriminatory power between mental health-positive and mental health-negative tweets.The precision score of **0.91** shows that ConvBiLSTM-AttnNet reliably identifies tweets indicative of mental health issues, minimizing false positives.With a recall of **0.87**, the model effectively identifies relevant mental health instances, ensuring minimal false negatives, which is crucial for accurate mental health detection.The F1 Score of **0.89** highlights the model's balanced performance, capturing both precision and recall strengths.The high AUC score of 96% underscores this model's ability to distinguish between classes, even at different decision thresholds. The ConvBiLSTM-AttnNet's performance can be attributed to its ability to capture both local features through convolutional layers and sequential dependencies with LSTM layers, while the attention mechanism focuses on contextually relevant parts of the tweet.



1. Receiver Operating Characteristic (ROC) Curve



1. Confussion Matrix for ConvBiLSTM-AttnNet

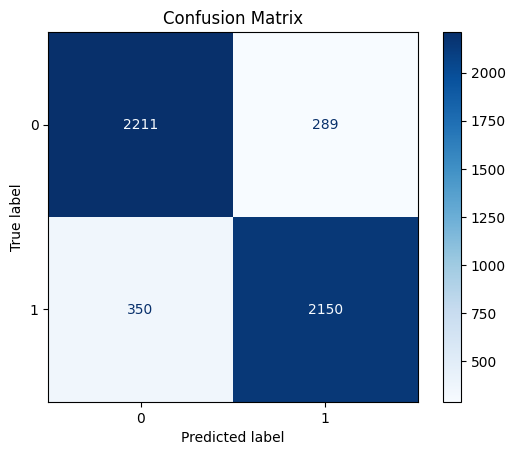


1. Training and Validation Loss Epochs
2. CLASSIFICATION REPORT

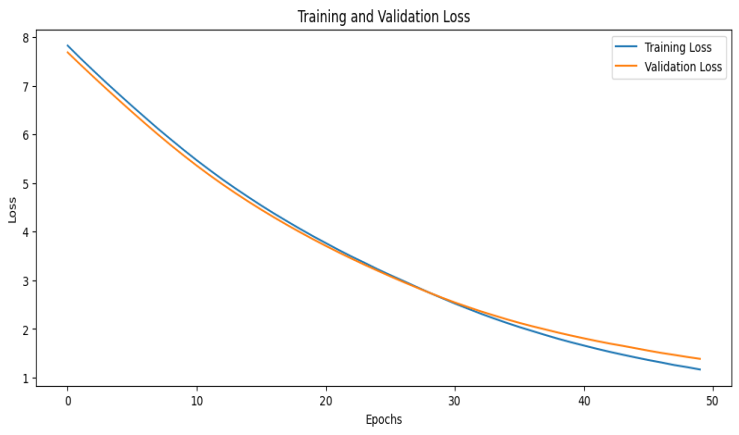
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | support |
| 0 | 0.88 | 0.91 | 0.89 | 2500 |
| 1 | 0.91 | 0.87 | 0.89 |  |
| accuracy |  |  | 0.89 |  |
| Mac avg | 0.89 | 0.89 | 0.89 | 5000 |
| Weighted avg | 0.89 | 0.89 | 0.89 | 5000 |

### *ConvLSTM-AttentionNet Results*

The ConvLSTM-AttentionNet model, which integrates Conv1D and LSTM layers along with an attention mechanism, also exhibited strong performance, achieving an **Accuracy of 87%** and an **AUC of 93%**. These metrics indicate that this model performs well in distinguishing between mental health-positive and negative tweets, albeit with slightly lower performance than ConvBiLSTM-AttnNet.The model’s precision of **0.88** signifies its ability to make accurate positive predictions, reducing the rate of false positives.With a recall of **0.86**, ConvLSTM-AttentionNet captures relevant positive cases effectively, though slightly less robustly than ConvBiLSTM-AttnNet. An F1 Score of **0.87** indicates that this model maintains a balanced performance, suitable for scenarios that require a trade-off between precision and recall.The slightly lower AUC score of 93% reflects a small drop in its ability to handle varied decision thresholds compared to ConvBiLSTM-AttnNet. Nevertheless, ConvLSTM-AttentionNet demonstrates solid performance, particularly in its interpretability due to the attention mechanism, which emphasizes critical parts of the sequence.



1. Confussion Matrix for ConvLSTM-AttentionNet



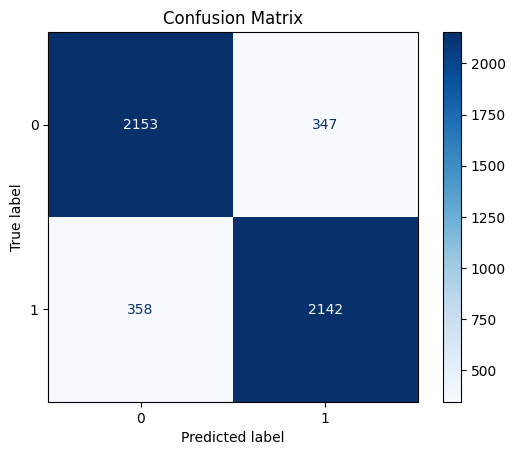
1. Training and Validation Loss Epochs
2. CLASSIFICATION REPORT:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | support |
| 0 | 0.86 | 0.88 | 0.87 | 2500 |
| 1 | 0.88 | 0.86 | 0.87 | 2500 |
| accuracy |  |  | 0.87 | 5000 |
| Mac avg | 0.87 | 0.87 | 0.87 | 5000 |
| Weighted avg | 0.87 | 0.87 | 0.87 | 5000 |

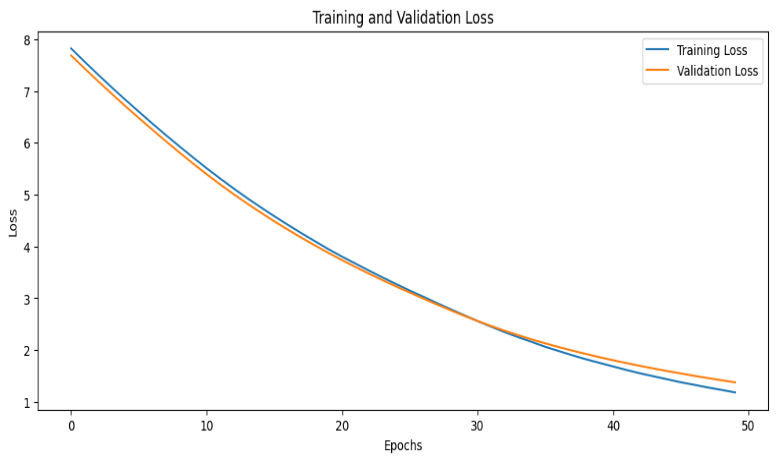
### 

### *HybridNet Results*

The HybridNet model, which employs L2-regularized Conv1D layers, bidirectional LSTM, and attention layers, achieved an **Accuracy of 86%** and an **AUC of 91%**. While it ranks slightly lower in performance than the previous models, HybridNet still provides reliable results in mental health classification.The precision score of **0.86** indicates a moderate ability to correctly identify positive cases, though it slightly lags behind the other models.With a recall of **0.86**, the model demonstrates adequate performance in identifying relevant instances.The F1 Score of **0.86** reflects a well-balanced but moderate performance across both precision and recall.The AUC score of 91% suggests that HybridNet can reasonably differentiate between positive and negative cases across different thresholds. The use of L2 regularization and a simpler attention mechanism makes HybridNet less prone to overfitting, though it slightly underperforms in comparison to ConvBiLSTM-AttnNet and ConvLSTM-AttentionNet.



1. Confusion Matrix of HybridNet Results



1. Training and Validation Loss Epochs
2. CLASSIFICATION REPORT:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | support |
| 0 | 0.86 | 0.86 | 0.86 | 2500 |
| 1 | 0.86 | 0.86 | 0.86 | 2500 |
| accuracy |  |  | 0.86 | 5000 |
| Mac avg | 0.86 | 0.86 | 0.86 | 5000 |
| Weighted avg | 0.86 | 0.86 | 0.86 | 5000 |

The ConvBiLSTM-AttnNet model emerged as the best performer among the three, achieving the highest accuracy, F1 Score, and AUC. Its architecture, which integrates convolutional and bidirectional LSTM layers with an attention mechanism, enables it to effectively capture both local features and long-term dependencies in the text, while the attention layer enhances focus on contextually important elements within the tweet. This combination of architectural elements explains its superior performance, especially in capturing the nuanced language associated with mental health topics on social media.ConvLSTM-AttentionNet, with a simpler architecture that excludes bidirectional LSTM, performed slightly below ConvBiLSTM-AttnNet. However, it achieved commendable accuracy and AUC scores, demonstrating that a streamlined architecture with attention mechanisms can still effectively classify mental health-related text. This model may be preferred in scenarios that prioritize computational efficiency without significantly compromising performance.HybridNet, while ranking the lowest in terms of accuracy and AUC, still provides reliable results with balanced performance metrics. Its use of L2 regularization contributes to model robustness, making it less susceptible to overfitting. HybridNet may be suitable for applications that require a simpler model with moderate performance and stability over different datasets.

All three models performed well in detecting depressive symptoms from Twitter data, with ConvBiLSTM-AttnNet showing the best overall performance. The integration of convolutional, sequential, and attention mechanisms across these models demonstrates the effectiveness of hybrid architectures in extracting meaningful patterns from social media text. The results validate our approach, highlighting the potential of deep learning models to support mental health monitoring based on real-time social media data.Each model’s performance across metrics demonstrates its suitability for different deployment scenarios, whether prioritizing high precision, balanced performance, or computational efficiency. These findings underscore the importance of model architecture in designing effective mental health classification systems and suggest promising directions for future research in applying hybrid deep learning techniques to unstructured text data.

**CHAPTER 5**

**CONCLUSION AND FUTURE WORK**

This study presented a comprehensive approach to detecting depressive indicators within Twitter data using advanced hybrid deep learning models. By combining convolutional, recurrent, and attention layers, our models effectively captured complex textual patterns associated with mental health symptoms, demonstrating that hybrid architectures can significantly improve mental health classification from social media data. The ConvBiLSTM-AttnNet model, in particular, emerged as the most robust, achieving the highest AUC and accuracy scores. These results underscore the value of leveraging both sequential dependencies and attention mechanisms in analyzing unstructured text data from social platforms. Our approach offers a scalable solution for mental health monitoring, potentially assisting clinicians and researchers in identifying at-risk individuals based on real-time social media data.

While this research has shown promising results, several areas could be explored to further advance the effectiveness of mental health classification from social media data:Future research could benefit from utilizing larger and more diverse datasets, covering a wider range of mental health conditions beyond depression. Incorporating data from multiple social media platforms, such as Reddit, Instagram, and Facebook, would enrich the dataset, offering broader insights into online mental health discourse. This expansion could help in developing models that generalize better across platforms and capture varying linguistic styles and emotive expressions.Although our study demonstrated the effectiveness of ConvBiLSTM-AttnNet and related models, future work could explore even more advanced architectures. Incorporating transformers or graph-based neural networks could enhance the model’s capacity to capture deeper context and semantic relationships within social media text. Additionally, combining models with sentiment and emotion recognition modules may improve detection accuracy by analyzing underlying emotional states more effectively.As social media language evolves rapidly, incorporating real-time adaptations for slang, abbreviations, and emerging expressions could further enhance model performance. Regularly updating N-gram libraries and employing adaptive embeddings, like domain-specific word embeddings, would make the models more resilient to changes in language patterns across social platforms.The ability to analyze mental health indicators in real-time could provide timely interventions for users showing signs of mental distress. Implementing a model capable of real-time analysis on social media streams would be invaluable for proactive mental health monitoring. Additionally, expanding this approach to handle multilingual data would enhance its applicability across different language groups and cultural contexts.Incorporating explainable AI (XAI) techniques into mental health classification models would enable greater transparency, making it easier for mental health professionals to interpret model outputs. Attention mechanisms and visual explainability tools, such as SHAP or LIME, could provide insights into the aspects of social media text that contribute most to predictions, increasing the model's utility in clinical settings.This study lays a strong foundation for mental health classification from social media data, demonstrating the potential of hybrid deep learning models. However, by advancing the dataset, model architecture, and language adaptability, future research can further enhance the capabilities and applicability of these models, ultimately contributing to more effective, scalable, and ethically responsible mental health monitoring solutions on social media platforms.

**CHAPTER 6**

**APPENDIX**

**Main Code**

from sklearn.metrics import roc\_curve, auc

import matplotlib.pyplot as plt

# Get model predictions as probabilities

y\_prob = model.predict(X\_test\_seq).ravel() # Flatten the probabilities for ROC curve

# Calculate the ROC curve and AUC score

fpr, tpr, thresholds = roc\_curve(y\_test, y\_prob)

roc\_auc = auc(fpr, tpr)

# Plot ROC curve

plt.figure(figsize=(8, 6))

plt.plot(fpr, tpr, color='blue', lw=2, label=f'ROC curve (AUC = {roc\_auc:.2f})')

plt.plot([0, 1], [0, 1], color='gray', linestyle='--') # Diagonal line for random chance

plt.xlabel('False Positive Rate')

plt.ylabel('True Positive Rate')

plt.title('Receiver Operating Characteristic (ROC) Curve')

plt.legend(loc="lower right")

plt.show()

**#ConvBiLSTM-AttentionNet & ConvLSTM-AttentionNet & HybridNet**

import tensorflow as tf

from tensorflow.keras.layers import Input, Embedding, Conv1D, GlobalMaxPooling1D, Bidirectional, LSTM, Dense, Dropout, BatchNormalization, Multiply, Activation

from tensorflow.keras.regularizers import l2

from tensorflow.keras.models import Model

from tensorflow.keras.callbacks import EarlyStopping

from tensorflow.keras.optimizers import AdamW

from tensorflow.keras.preprocessing.text import Tokenizer

from tensorflow.keras.preprocessing import sequence

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import classification\_report, confusion\_matrix, ConfusionMatrixDisplay

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

df = pd.read\_csv('/content/Mental-Health-Twitter.csv')

# Prepare the data

X = df['post\_text']

y = df['label']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, stratify=y, random\_state=42)

# Tokenize and pad sequences

tok = Tokenizer()

tok.fit\_on\_texts(X\_train)

train\_sequences = tok.texts\_to\_sequences(X\_train)

test\_sequences = tok.texts\_to\_sequences(X\_test)

vocab\_size = len(tok.word\_index) + 1

MAX\_LEN = 40

X\_train\_seq = sequence.pad\_sequences(train\_sequences, maxlen=MAX\_LEN)

X\_test\_seq = sequence.pad\_sequences(test\_sequences, maxlen=MAX\_LEN)

# Define attention mechanism

def attention\_block(inputs):

attention = Dense(1, activation='tanh')(inputs)

attention = tf.keras.layers.Flatten()(attention)

attention = Activation('softmax')(attention)

attention = tf.keras.layers.RepeatVector(inputs.shape[-1])(attention)

attention = tf.keras.layers.Permute([2, 1])(attention)

output = Multiply()([inputs, attention])

return output

# Define the hybrid model architecture with L2 regularization and higher dropout

inputs = Input(shape=(MAX\_LEN,), name="inputs")

x = Embedding(input\_dim=vocab\_size, output\_dim=128)(inputs)

# Convolutional Block

x = Conv1D(filters=64, kernel\_size=3, activation='relu', padding='same', kernel\_regularizer=l2(0.01))(x)

x = BatchNormalization()(x)

x = Conv1D(filters=128, kernel\_size=3, activation='relu', padding='same', kernel\_regularizer=l2(0.01))(x)

x = GlobalMaxPooling1D()(x)

# LSTM Block with Attention

x = tf.keras.layers.Reshape((1, -1))(x) # Reshape for attention compatibility

x = Bidirectional(LSTM(64, return\_sequences=True, kernel\_regularizer=l2(0.01)))(x)

x = attention\_block(x)

x = tf.keras.layers.Flatten()(x)

# Fully connected layers with increased dropout

x = Dense(128, activation='relu', kernel\_regularizer=l2(0.01))(x)

x = Dropout(0.6)(x) # Increased dropout

x = Dense(64, activation='relu', kernel\_regularizer=l2(0.01))(x)

x = Dropout(0.6)(x) # Increased dropout

outputs = Dense(1, activation='sigmoid')(x)

# Model creation and compilation

model = Model(inputs, outputs)

optimizer = AdamW(learning\_rate=1e-5) # Reduced learning rate

model.compile(loss='binary\_crossentropy', optimizer=optimizer, metrics=['accuracy'])

# Model summary

model.summary()

# Define early stopping with increased patience

early\_stopping = EarlyStopping(monitor='val\_loss', patience=10, restore\_best\_weights=True)

# Train the model

history = model.fit(X\_train\_seq, y\_train,

epochs=100, # Increased epochs to give more time for improvement

validation\_split=0.2,

batch\_size=64,

callbacks=[early\_stopping])

# Evaluate the model

test\_loss, test\_accuracy = model.evaluate(X\_test\_seq, y\_test)

print(f'\nTest Loss: {test\_loss}, Test Accuracy: {test\_accuracy}')

# Predictions and classification report

y\_pred = (model.predict(X\_test\_seq) > 0.5).astype("int32")

print("\nClassification Report:")

print(classification\_report(y\_test, y\_pred))

# Confusion matrix

cm = confusion\_matrix(y\_test, y\_pred)

ConfusionMatrixDisplay(cm).plot(cmap="Blues")

plt.title('Confusion Matrix')

plt.show()

# Plot training and validation accuracy and loss

history\_dict = history.history

# Accuracy plot

plt.figure(figsize=(12, 5))

plt.plot(history\_dict['accuracy'], label='Training Accuracy')

plt.plot(history\_dict['val\_accuracy'], label='Validation Accuracy')

plt.xlabel('Epochs')

plt.ylabel('Accuracy')

plt.title('Training and Validation Accuracy')

plt.legend()

plt.show()

# Loss plot

plt.figure(figsize=(12, 5))

plt.plot(history\_dict['loss'], label='Training Loss')

plt.plot(history\_dict['val\_loss'], label='Validation Loss')

plt.xlabel('Epochs')

plt.ylabel('Loss')

plt.title('Training and Validation Loss')

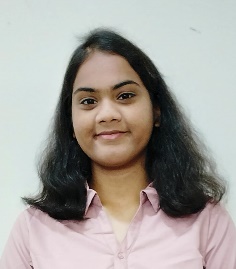
plt.legend()

plt.show()

**REFERENCES**

1. L. Liu, "Research on Logistic Regression Algorithm of Breast Cancer Diagnose Data by Machine Learning," *2018 International Conference on Robots & Intelligent System (ICRIS)*, Changsha, China, 2018, pp. 157-160, doi: 10.1109/ICRIS.2018.00049.
2. Brandon A. Kohrt1,2, Nagendra P. Luitel1\* Abstract, Prakash Acharya1 and Mark J. D. Jordans1,3,4
3. Kailai Yanga, Tianlin Zhanga, Sophia Ananiadoua,b,
4. Learning Riya Aggarwal 1Computer Science and Engineering, Amity University Noida, Uttar Pradesh, India 1riyaagg09@gmail.com Anjali Goyal2 2 School of Engineering and Technology, Sharda University Greater Noida, Uttar Pradesh, India
5. A. Deshpande and M. Kappali, "Coordination of Energy Storage Systems and Closed Loop Converter for Regulating Voltage of PV System," *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, 2020, pp. 1-5, doi:10.1109/CONECCT50063.2020.9198444.
6. Genaro K, Fabris D, Arantes ALF, Zuardi AW, Crippa JAS, Prado WA. Cannabidiol Is a Potential Therapeutic for the Affective-Motivational Dimension of Incision Pain in Rats. Front Pharmacol. 2017 Jun 21;8:391. doi: 10.3389/fphar.2017.00391. PMID: 28680401; PMCID: PMC5478794.
7. Chen X, Li X, Zhang JJ, Han Y, Kan J, Chen L, Qiu C, Santoso T, Paiboon C, Kwan TW, Sheiban I, Leon MB, Stone GW, Chen SL; DKCRUSH-V Investigators. 3-Year Outcomes of the DKCRUSH-V Trial Comparing DK Crush With Provisional Stenting for Left Main Bifurcation Lesions. JACC Cardiovasc Interv. 2019 Oct 14;12(19):1927-1937. doi: 10.1016/j.jcin.2019.04.056. Epub 2019 Sep 11. PMID: 31521645.
8. †Hong Chen and Songhua Hu contributed equally to this work and should be considered co-first authors. 3School of Mathematic and Statistic, Hubei University of Science and Technology, Xianning, China Full list of author information is available at the end of the article
9. R. MurtiRawat, S. Panchal, V. K. Singh and Y. Panchal, "Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning," *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2020, pp. 534-540, doi: 10.1109/ICESC48915.2020.9155783.
10. T. Hossain, F. S. Shishir, M. Ashraf, M. A. Al Nasim and F. Muhammad Shah, "Brain Tumor Detection Using Convolutional Neural Network," *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934561.
11. N. Klugman, J. Vedral and J. Lang, "Field-based Model of Flux Compression Generators," *2020 International Applied Computational Electromagnetics Society Symposium (ACES)*, Monterey, CA, USA, 2020, pp. 1-2, doi: 10.23919/ACES49320.2020.9196197.
12. Van den Berg JD, Quintens L, Zhan Y, Hoekstra H. Why address posterior tibial plateau fractures? Injury. 2020 Dec;51(12):2779-2785. doi: 10.1016/j.injury.2020.09.011. Epub 2020 Sep 16. PMID: 32958346.
13. Ting L, Yan-Hong L, Song Z, Chun-Rong X, Xuan D, Jian-Feng Z, Wei L, Qing-Jie Y, Kun Y. [Rapid detection of *Schistosoma japonicum*-infected snails with recombinase-aided isothermal amplification assay]. Zhongguo Xue Xi Chong Bing Fang Zhi Za Zhi. 2019 Apr 24;31(2):109-114. Chinese. doi: 10.16250/j.32.1374.2019026. PMID: 31184038.
14. M. R. Machado, S. Karray and I. T. de Sousa, "LightGBM: an Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry," *2019 14th International Conference on Computer Science & Education (ICCSE)*, Toronto, ON, Canada, 2019, pp. 1111-1116, doi: 10.1109/ICCSE.2019.8845529.
15. https://github.com/Microsoft/LightGBM 2http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/gbm.html 3https://github.com/catboost/catboost
16. KDD’18, August 19–23, 2018, London, United Kingdom ©2018 Copyright held by the owner/author(s). Publication rights licensed to Associa tion for Computing Machinery. ACMISBN978-1-4503-5552-0/18/08...$15.00
17. S. Sehgal, H. Singh, M. Agarwal, V. Bhasker and Shantanu, "Data analysis using principal component analysis," *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*, Greater Noida, India, 2014, pp. 45-48, doi: 10.1109/MedCom.2014.7005973.
18. D. Tomar, Y. Prasad, M. K. Thakur and K. K. Biswas, "Feature Selection Using Autoencoders," *2017 International Conference on Machine Learning and Data Science (MLDS)*, Noida, India, 2017, pp. 56-60, doi: 10.1109/MLDS.2017.20.
19. Rai NK, Rim KI, Wulandari EW, Subrata F, Sugihantono A, Sitohang V. Strengthening emergency preparedness and response systems: experience from Indonesia. WHO South East Asia J Public Health. 2020 Apr;9(1):26-31. doi: 10.4103/2224-3151.282992. PMID: 32341218.
20. Y. Zhang, X. Li, L. Pang, Y. He, G. Ren and J. Li, "A 2-D Geometry-Based Stochastic Channel Model for 5G Massive MIMO Communications in Real Propagation Environments," in *IEEE Systems Journal*, vol. 15, no. 1, pp. 307-318, March 2021, doi: 10.1109/JSYST.2020.2971062.
21. Y. Li, R. Wang, H. Liu, H. Jiang, S. Shan and X. Chen, "Two Birds, One Stone: Jointly Learning Binary Code for Large-Scale Face Image Retrieval and Attributes Prediction," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 3819-3827, doi: 10.1109/ICCV.2015.435.
22. Zhang C, Kong Y, Shen K. The Age, Sex, and Geographical Distribution of Self-Reported Asthma Triggers on Children With Asthma in China. Front Pediatr. 2021 Sep 3;9:689024. doi: 10.3389/fped.2021.689024. PMID: 34540763; PMCID: PMC8448385.
23. S. Zade, R. Korde, R. Sonone and M. Shah, "Performance Analysis of Parallel Image Processing Operations," *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2020, pp. 0597-0601, doi: 10.1109/ICCSP48568.2020.9182069.
24. A. -r. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. E. Hinton and M. A. Picheny, "Deep Belief Networks using discriminative features for phone recognition," *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011, pp. 5060-5063, doi: 10.1109/ICASSP.2011.5947494.
25. Brands MT, Verbeek ALM, Geurts SME, Merkx MAW. Gepersonaliseerde nacontrole bij patiënten die curatief zijn behandeld voor een plaveiselcelcarcinoom in de mondholte [Personalised follow-ups for patients curatively treated for squamous cell carcinoma in the oral cavity]. Ned Tijdschr Tandheelkd. 2021 Mar;128(3):167-172. Dutch. doi: 10.5177/ntvt.2021.03.20112. PMID: 33734223.
26. Taylor B, Cleary S. A retrospective, observational study of medicolegal cases against obstetricians and gynaecologists in South Africa's private sector. S Afr Med J. 2021 Jun 30;111(7):661-667. doi: 10.7196/SAMJ.2021.v111i7.15511. PMID: 34382550.
27. V. K. V and P. Samuel, "A Multinomial Naïve Bayes Classifier for identifying Actors and Use Cases from Software Requirement Specification documents," *2022 2nd International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 2022, pp. 1-5, doi: 10.1109/CONIT55038.2022.9848290.
28. Genaro K, Fabris D, Arantes ALF, Zuardi AW, Crippa JAS and Prado WA (2017) Cannabidiol Is a Potential Therapeutic for the Affective-Motivational Dimension of Incision Pain in Rats. Front. Pharmacol. 8:391. doi: 10.3389/fphar.2017.00391
29. A. Zafar and S. Chitnis, "Survey of Depression Detection using Social Networking Sites via Data Mining," *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2020, pp. 88-93, doi: 10.1109/Confluence47617.2020.9058189.
30. Chen X, Li X, Zhang JJ, Han Y, Kan J, Chen L, Qiu C, Santoso T, Paiboon C, Kwan TW, Sheiban I, Leon MB, Stone GW, Chen SL; DKCRUSH-V Investigators. 3-Year Outcomes of the DKCRUSH-V Trial Comparing DK Crush With Provisional Stenting for Left Main Bifurcation Lesions. JACC Cardiovasc Interv. 2019 Oct 14;12(19):1927-1937. doi: 10.1016/j.jcin.2019.04.056. Epub 2019 Sep 11. PMID: 315216

**BIODATA**



Name : Abhilipsa Padhy

Mobile Number : 9439801304

E-mail : abhilipsa.21bce8399 @vitap.ac.in

Permanaent Address : Quator no.4R/14,Unit-8, Bhubaneswar, Odisha

****

Name : Ishan Das

Mobile Number : 6202122496

E-mail : ishan.21bce7578 @vitap.ac.in

Permanaent Address : F4/4 marine drive road, Jamshedpur, Jharkhand