I.J. Information Technology and Computer Science, **2024, 4, 45-55**
Published Online on August 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2024.04.03

An Optimized Convolutional Neural Network Model for Detecting Depressive Symptoms from Image Posts

Awoyelu T. M.*
Department of Computer Sciences, Faculty of Natural Sciences, Ajayi Crowther University, Oyo, Nigeria
E-mail: tm.awoyelu@acu.edu.ng
ORCID iD: https://orcid.org/0000-0002-8390-8385
*Corresponding author

Iyanda A. R.
Department of Computer Science and Engineering, Faculty of Technology, Obafemi Awolowo University, Ile-Ife, Nigeria
E-mail: abiyanda@oauife.edu.ng
ORCID iD: https://orcid.org/0000-0002-3470-0115

Mosaku S. K.
Department of Mental Health, College of Health Sciences, Obafemi Awolowo University, Ile-Ife, Nigeria
E-mail: kmosaku@yahoo.co.uk, skkolawole@gmail.com

Received: 18 September 2023; Revised: 20 November 2023; Accepted: 26 January 2024; Published: 08 August 2024

Abstract: This paper presents an optimized model that uses an optimized CNN to detect depressive symptoms from image posts. This is with a view to detecting depression symptoms in individuals. Visual data were collected in their raw form and assessed as having or not having a mental condition. The images were processed, and the relevant features retrieved from them. An optimized convolutional neural network (CNN) was used to simulate the defined classification model of the image posts. The model was implemented using Python Programming Language. Precision, recall, accuracy, and the area under the Receiver Operating Characteristics (ROC) curve were used as performance indicators to assess the model's efficacy. The collected findings indicate that 77% accuracy is achieved by the optimized model. As a result, 77% of the cases were accurately predicted by the model, suggesting that the model is generally accurate in its predictions. The research will contribute to a decrease in the incidence, prevalence, and recurrence of mental health illnesses as well as the disabilities they cause.

Index Terms: Depression, Images, Convolutional Neural Network, Twitter, Accuracy.

1. Introduction

Healthcare workers are increasingly using smartphone selfies and photographs to acquire images of injuries, infections, skin wounds, and the eyes. This use of smart technology allows doctors to analyze a patient without needing to be there in person, minimizing the time a patient must wait for a diagnosis and combating the scarcity of specialists [1]. It is also beneficial in distant places or impoverished countries when healthcare is limited and difficult to access. In recent studies, machine learning (ML) has grown in popularity. Image identification, multimedia concept retrieval, social network analysis, video recommendation, text mining, and other applications have all used it. In these applications, deep learning (DL) is commonly employed [2]. Deep learning is a group of techniques that uses a multi-layer neural network to handle problems like images and text using a range of machine learning algorithms. Deep learning is capable of yielding more intellectual advanced feature classes when paired with low-level characteristics, allowing data with dispersed feature representations to be discovered. Due to the expansion of deep learning, Multimodal Deep Learning (MMDL) has recently attracted the interest of scientists. People see, hear, touch, smell, and taste objects around them, and they have a multimodal experience of them. Multiple elements of items are collected in order to convey data in various media formats such as picture, text, video, graph, sound and so on. Modality refers to the representation format in which a specific sort of data is kept. As a result, the various media types listed above correspond to modalities, and the representation of these numerous modalities as a whole is referred to as multimodal [3]. Multimodal learning techniques create a model by mixing several data types, such as audio, text, photos, video, and lot more, to tackle problems ranging from data alignment to categorization from multiple sources [4]. Physical and mental well-being remain elaborately intertwined. The mental health of people has a great impact on their ability to sustain good physical health. Individuals suffering from mental diseases, like anxiety, fear and sorrow, find it difficult to be involved in health-promoting activities. Chronic illnesses, for example, can have a devastating effect on a person's mental health and make it challenging to participate in treatment and rehabilitation [5]. According to [6], depressive symptoms can show through both verbal and nonverbal communication, using both texts and photos, which are the most prevalent types of user-generated data.

Detecting depressed symptoms is a challenging endeavor that involves assessing multiple sorts of data, including textual, visual, and behavioral indicators. Scholars have investigated many methods to pinpoint these signs. These include machine learning and deep learning, text-based approaches, multimodal approaches, natural language processing (NLP), behavioral and biometric data, social media network analysis, questionnaires and self-reports, and cross-domain approaches. These methods can be used separately or in combination. The creation of precise and trustworthy techniques for identifying depression symptoms is essential for early intervention and helping those who require support. Numerous methods for identifying depressive symptoms have produced noteworthy findings, furthering the field of mental health research. According to study, sentiment analysis of text can successfully identify depressed symptoms for text-based approaches. It has been demonstrated that combining data from several domains—such as text, photos, and behavior—can enhance the precision and resilience of models used to identify depression symptoms. In the topic of mental health informatics, the majority of recent research has been on methods for detecting depressed symptoms that are either text- or image-based.

Major mental illnesses including affective bipolar illness, dementia, and depression are reported to be among the top ten causes of disability in adults worldwide. In contrast, the majority of adult mental health issues originate during childhood and teenage years. The global rate of impairment, which is primarily brought on by mental illnesses, will decline with early detection of mental disorders.

The various approaches used in detecting depressive symptoms from multimodal data sources using hand-rafted features and baseline CCN model have given low accuracy. While previous studies have explored the use of various modalities such as audio, video and text for depressive symptoms, relatively little research has been done on the use of images and textual posts. To address the highlighted problems, there is a need for an optimized model that will not only offer the automatic feature extraction, but will also improve the classification accuracy through the use of CNN. Hence, this paper aims to develop a model that uses images for the prediction of depression symptoms. Mental disorder prevention will help to reduce the occurrence, prevalence, and recurrence of mental health disorders and their associated disability. This will reduce suicide rate, adolescent suicide attempts, the number of people who engage in disordered eating pattern in an effort to manage their weight, and the number of persons who experience Major Depressive Episodes (MDEs).

2. Related Works

[7] worked on predicting depression via social media. Crowdsourcing was used to gather twitter profiles and extract tweets from users' publicly accessible profiles. SVM classifier was built that can predict an individual's likelihood of depression ahead of the reported onset of depression. The result obtained gave 70% classification accuracy. Using a common psychometric tool and crowdsourcing, the study assembled a group of Twitter users with clinical depression diagnoses; it made no attempt to search other social media sites.

[8] aimed to identify depression using Twitter activity. SVM was used in machine learning to create classifiers that estimate the existence of active depression. A 10-fold cross-validation procedure was used to assess the categorization accuracy. The classifiers were assessed using precision, recall, accuracy of the estimations, and F-measure. It was discovered that assigning a score of 0, 0, 0, and 0.61 to each of the precision, recall, F-measure, and accuracy for each participant who was classified as not having depression. The comparable numbers in the complementary situation are 0.39, 1, 0.56, and 0.39, respectively, when all subjects are categorized as having depression. It is anticipated that the newest machine learning frameworks, like deep learning and ensemble learning, will be used instead of just SVM. This could lead to an even higher estimation accuracy. The study used information from only 209 users.

[9] sought to identify users who were at danger by tracking tweets for signs of depression. 122 million tweets were gathered from Twitter. The model was implemented in R, and the dataset was trained using linear SVM. The recall value of the system was 85%, and its accuracy was 78.72%. The study solely used data from Twitter and considered textual posts. [10] looked at predicting the likelihood of depression using text, image and behavior data from Instagram. Data was collected by crawling Instagram posts containing hashtags related to depression for depressed users and the word "happy" for non-depressed users. There were 512 users in total. CNN was utilized for picture modeling and Word2vec for text modeling. The model was assessed using Precision, recall and F1-score. The system's recall value was 73.5%, its precision was 88.8%, and its area under the ROC was greater than 0.5. In addition to outperforming the method using handcrafted features, the strategy of automatically learning features from photos and creating a deep learning prediction model also saves labor costs. According to the study, users' photos exhibit signs of despair, and modeling textual and visual aspects jointly outperforms modeling them separately. 512 people were employed for the study, despite the fact that the work included both texts and graphics. Furthermore, word2vec is not appropriate for use where text classification is needed. Word2Vec's incapacity to handle unfamiliar terms is a problem, especially in domains like Twitter where a lot

of sparse and noisy data may have been utilized.

[11] worked on social media multimodal mental health analysis. The study provided a thorough examination of the contextual and visual content of Twitter profiles that were probably depressed. Statistical techniques were used to combine disparate sets of information gained from analyzing textual, visual, and user interactions into a multimodal framework. Based on an empirical study, the multimodal framework was found to be superior, as evidenced by the 5% improvement in the average F1-Score. The study sheds light on the connection between mental health and demography.

[12] worked on utilizing a hybrid deep learning model to identify depression in Reddit posts. Using the PRAW Reddit API, textual data were gathered from Reddit (2,000,000 posts). The work combined CNN and LSTM in a hybrid deep learning method to detect depression in user posts. The model's accuracy, recall, and precision were assessed. All three models—CNN, LSTM, and LSTM+CNN—reported accuracy values of 92.60%, 92.62%, and 93.98%, respectively. Only textual posts were employed in the study. Investigating various modalities and social media network pairings can improve the model's effectiveness. [13] worked on a hybrid learning strategy that combined feature-rich CNN and bi-directional LSTM to predict depression from user tweets. 292,564 tweets from 1402 individuals on Twitter were gathered. The model's accuracy, recall, and precision were assessed. Higher accuracy, precision, f1-score, specificity, and 96.28, 96.99, 94.78, and 96.35 were obtained by the hybrid CNN-biLSTM model. Investigating various modalities and machine learning algorithm combinations can improve the accuracy of the model. Several machine learning techniques can be assessed on diverse social media platforms in addition to Twitter raw data. [14] developed a predictive model to determine whether university students suffer from depression. The model's framework is depicted in Fig. 1, which also demonstrates how the dataset—historical data—was collected from students using a questionnaire and pre-processed. The most pertinent risk variables for depression were then identified by applying a genetic algorithm to the processed data. Genetic programming was then utilized to create a depression prediction model based on the identified risk factors. Accuracy, true positive rate, precision, and false positive rate were used to assess the model's performance. Based on the results of the simulation, 465 out of 507 records were successfully identified by the model without feature selection, with an accuracy of 91.7%, whereas 475 out of 507 records were correctly classified by the model with feature selection, with an accuracy of 93.7%. The predictive model yielded 422 correct classifications with an accuracy of 83.23% for the five-class dataset, both with and without feature selection. The methodology used manually completed questionnaires to predict depression. A significant obstacle associated with this methodology is that the system's functionality is contingent upon the data given by the involved parties. On the other hand, it was possible to give false information.

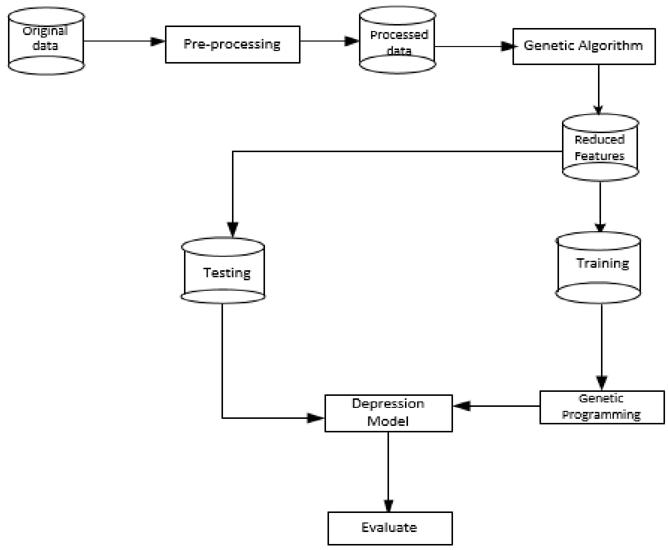


Fig.1. Existing model [14]

This paper fills an important gap in current research by exploring the potential of images posts to predict depression severity using CNN. By doing so, it could provide valuable insights into how CNN models can be used to improve early detection and intervention in mental health disorders.

3. Methodology

The experimental process to achieve the objectives of this study are: data collection, data pre-processing, dataset splitting into training and testing set, model formulation was done using CNN and Adam optimization.

First, unprocessed visual data were gathered, and then their mental health status was determined. After processing, the images' pertinent features were extracted. The specified categorization model of the photo postings was simulated using an efficient convolution neural network (CNN). Using the collected historical data, the system's performance was assessed in terms of accuracy, precision, recall, and the area under the Receiver Operating Characteristics (ROC) curve.

Experts tested the method with Mean Opinion Score.

3.1. Data Collection

The Twitter dataset contains 1.6million posts, made up of 800,000 non-depressive posts. The preprocessed images acquired contained 28,709 images, 13,358 of which are depressive and the remaining 15,351 are non-depressive images. The training set, validation set, and testing set of this dataset were split up into proportions of 70%, 15%, and 15%, respectively. The breakdown of data is as Table 1 illustrates.

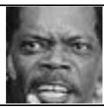
Table 1. Image data division

Image Data Division	Number of Images
Total depressed images	13358
Total undepressed images	15351
Training depressed images	9350
Training undepressed images	10745
Validation depressed images	2003
Validation undepressed images:	2302
Test depressed images	2005
Test undepressed images	2304
Total training sample	20095 (percentage:70%)
Total validation sample	4305 (percentage:15%)
Total testing sample	4309 (percentage:15%)

Table 2. Sample dataset of non -depressive posts

S/N	Username	Textual Post	Image Post
2	becca210	Didn't place in the Peeps contest but thanks for voting anyways.	
		Going to bed so goodnight, everyone and sweet dreams http://twitpic.com/2y2e0	
		@LittleLumen walking over to put the deposit down tomorrow	
		is now followinq @DAChesterFrench, you shud do tha same	
		@LordPov Are you meant to add on the back of that "twittering from a toilet cubicle somewhere"?	
		Aw i'm holding my new puppy. Well, He's not mine but He's a cutie.	
		@iJohn kitteh is sleepin on my crotch which proves she likes me more	

Table 3. Sample data of depressive posts

S/N	Username	Textual Post	Image Post
1	_fuurbare	Hbl makes me wanna hang myself to death	
		Hey everyone! If you like seeing the best in-game deaths (XD) or just want to hang out with a crazy bear like myself, come join me at my @Twitch channel YeetMcFleek. See you there! #HelloThere #JoinTheJollity #NotMLG #twitchstreamer #NotLikeThis pic.twitter.com/E3g3wIiuGg	
		After leaving from Iraq, we left knowing we were the bad guys. And that fighting for freedom was a farce. All my friends that I used to hang out with died horrible deaths. Life hasn't been the same. At least the VA is there to keep me from killing myself.	
		Hang on there, brah. Iâ€™m 65 myself. Iâ€™m dedicated to bringing down the orange scourge & his motley crew. Iâ€™m all in. Itâ€™s victory -in November- or death - American democracy Letâ€™s get rid of the republican vermin. Thatâ€™s a great 1st step. YESSS	
		One that comes to mind is Blur's Death Of A Party. For years I thought he sang: "Go to another party and hang myself - Jelly on the shelf" When it's actually "Gently on the shelf" ðŸ™„	
		FAILURE FIIINNDD MEE TO TIE ME UP NOW CUZ IM AS BAD AS BAD AS IT GEEEETTSS FAILURRREE FIND MEEE \TO HANG ME UP (?) BY MY NECK CUZZ IM A FATE WORSE THAN DEATH what a CYAnide surprise you have left for my eyes If I had common sense I'd cut myself or curl up and die âœŒï¸□ðŸ˜Œ	
		Welp Off to hang myself. pic.twitter.com/xP4F6FRkC6	

The dataset's attributes include the user id - which uniquely identifies each user, the query information present in the dataset, the user id for the tweet - which indicates the unique number assigned to the tweet, the tweet's text. A sample dataset of non-depressive posts is shown in Table 2 and a sample dataset with depressive posts is shown in Table 3.

The dataset contains the following five (5) fields, namely:

- *ids* renamed SNo. This is the id of the post (1467811184).
- date renamed PostDate. This is the date of the post (Mon Apr 06 22:19:57 PDT 2009).
- flag renamed Flag. This is the query (lxy). If there is no query, then this value is NO QUERY.
- user renamed User. This is the user that posted (ElleCTF).
- Text renamed Post. This is the text of the post (my whole body feels itchy and like it's on fire).

3.2. Image Posts Preprocessing

Dimension of input: The input photographs have a reduced dimension format of (48, 48). The small size, however, prevents one from being able to investigate more deeply while creating a CNN model. Therefore, the first step is to convert the photos into a dimension length that is fixed (128, 128, 3).

Image Dataset Splitting into Training and Testing Set

The preprocessed images obtained contained 28,709 images; 13,358 of which are depressive and the remaining 15,351 are non-depressive images. A 70% training set, 15% validation set, and 15% testing set component were separated out of these datasets.

3.3. Model Formulation

For image modeling, the following existing techniques are to be used: CNN and Adam optimization.

3.4. Building the Image Model

In order to train the CNN model to identify whether or not an image is depressed, the preprocessed image is fed into the network. The configuration/parameter of the CNN model is as shown in Table 4 and the configuration for the image modelling is:

(Conv2D(32,(64, 64, 1),(5, 5)) + ReLU) ∗ 2 + M axP ool2D((2, 2), Stride(2, 2)) + (Conv2D(64,(32, 32, 1),(3, 3)) + ReLU) ∗ 2 + M axP ool2D(2, Stride(2, 2))+F C(256)+ Sigmoid(2).

Fig. 2 shows the architecture of the CNN model. The architecture shows the various stages the input image has to go through to detect whether an image is depressed or not.

Table 4. Configuration/parameter of the CNN model

Model 1: Sequential

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 44, 44, 32)	2432
conv2d_1 (Conv2D)	(None, 40, 40, 32)	25632
max_pooling2d (MaxPooling2D)	(None, 20, 20, 32)	0
conv2d_2 (Conv2D)	(None, 18, 18, 64)	18496
conv2d_3 (Conv2D)	(None, 16, 16, 64)	36928
max_pooling2d_1(MaxPooling2D)	(None, 8, 8, 64)	0
flatten (Flatten)	(None, 4096)	0
dense (Dense)	(None, 256)	1048832
dense_1 (Dense)	(None, 128)	32896
dense_2 (Dense)	(None, 64)	8256
dense_3 (Dense)	(None, 32)	2080
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 2)	34

Total params: 1,176,114
Trainable params: 1,176,114
Non-trainable params: 0

DeoCSC19-20

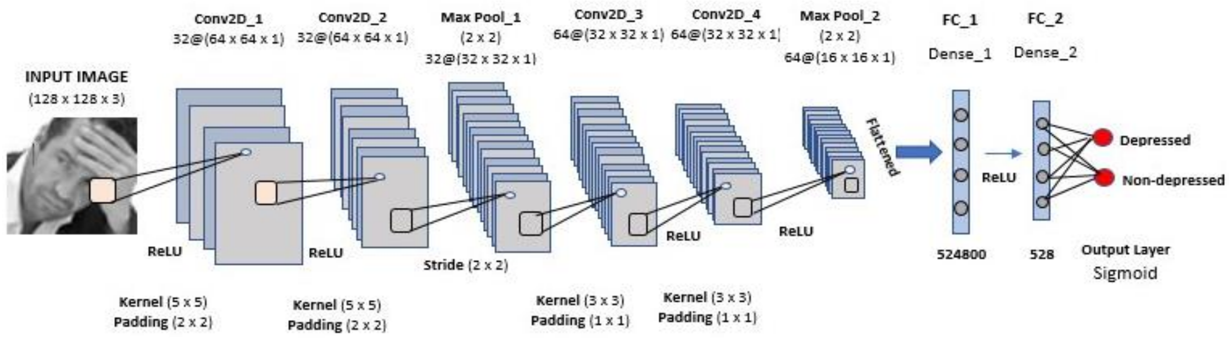


Fig.2. The Architecture of the CNN model [15]

The architecture has the following layers: Conv2D, MaxPooling2D, Flatten, Dense and Output Layers.

Conv2D Layers: These layers perform convolution operations on the input images. The first Conv2D layer has 32 filters with a kernel size of (5,5), and subsequent layers also have 32 and 64 filters with smaller kernel sizes. The output shape after convolution is determined by the filter count and kernel size.

MaxPooling2D Layers: After each pair of Conv2D layers, MaxPooling2D layers are added. These layers reduce spatial dimensions and help retain essential information while reducing computational complexity.

Flatten Layer: The output is transformed into a 1D vector by the Flatten layer after convolution and pooling, getting it ready for the fully linked layers.

Dense Layers: The series of Dense layers act as a traditional neural network for classification. The number of units in these layers gradually decreases, leading to the final output layer with 2 units (for depressed and non-depressed classes).

Output Layer: The final layer has 2 units with SoftMax activation, which generates class probabilities for prediction.

The 'Param Number' column represents the number of learnable parameters in each layer. These parameters are tuned during training to optimize the model's performance.

Conv2D Layers use convolutional filters to detect features in the input images. The size of the applied kernel and the number of filters define the shape of the output. The output shape of the first Conv2D layer, which contains 2,432 parameters, is (None, 44, 44, 32). The feature maps are down sampled using max pooling using the MaxPooling2D Layers. They have an output form, but as pooling entails choosing the highest value within a range, they need parameters. In order to prepare the feature maps for the fully connected layers, the Flatten Layer reshapes them into a 1D vector.

Dense Layers: The size of the applied kernel and the number of filters define the shape of the output. The output shape of the first Conv2D layer, which contains 2,432 parameters, is (None, 44, 44, 32). The feature maps are down sampled using max pooling using the MaxPooling2D Layers. They have an output form, but as pooling entails choosing the highest value within a range, they need parameters. In order to prepare the feature maps for the fully connected layers, the Flatten Layer reshapes them into a 1D vector.

3.5. CNN Model Optimization

The ideal network parameters were obtained using the Adam optimizer, which has continuously been demonstrated in the literature to produce a superior optimal solution. The Adam optimizer was configured using the Keras framework's default parameters. These default settings are the best that can be found, according to several examinations.

Training the model 28,709 preprocessed images were obtained, 13,358 of which show depression and 15,351 of which do not. Seventy percent, fifteen percent, and fifteen percent of the total are the training, validation, and testing sets of these datasets. The CNN model is trained using Adam optimization techniques on both the training and validation sets. A CNN model must go through two main stages of training. Both the training and convolution stages of a fully linked artificial neural network in the convolutions process, Max Pooling and convolutions are regularly used to recover unique characteristics (feature maps) of a certain image category.

This training consists of two steps: forward propagation and reverse propagation. In the forward propagation step (0), weights are equally started with a few random weight values that are slightly above zero. The activation function looks through the input values using these weights until the output class is identified. This function determines whether a node transmits its value to the network's subsequent tier. The backward propagation step in a neural network adjusts the weights based on the actual loss. Calculated is the difference, in this example equal to a loss, between the production that was expected and what was actually generated. Using the computed loss, the weights of the previous layers are adjusted based on how much they affected the outcome. Iterating through this procedure several times, or epochs, yields the optimal weights. The CNN model is trained over 25 epochs using the architecture described above. Every layer applied non-linearity using a ReLU activation function, with the exception of the output layer, which used a Sigmoid activation function (used for binary classification tasks). The "binary cross entropy" loss function is used to compute the loss, while "accuracy" serves as the performance metric. In order to facilitate a faster and more accurate convergence of the optimizer to the global minimum of the loss function (LR), the learning rate is annealed. The optimizer sets out on its journey over the "loss landscape" with the LR. As the LR increases, the steps get bigger and the convergence happens faster. Nevertheless, the optimizer may reach a high LR local minima that is misinterpreted as representing optimal performance.

A lower learning rate is desirable in order to efficiently approach the global minimum of the loss function during training. To preserve the advantages of a high LR and quick calculation time, the LR will be dynamically reduced every X step (epochs), depending on whether it is necessary (when accuracy is not improved after specified steps). If the accuracy is not improved after three epochs, the ReduceLROnPlateau function from Keras.callbacks is used to determine that the LR should be cut in half. More significantly, an early-stopping mechanism is used to halt the learning if the validation-loss is not decreased after 5 rounds, at which point it can be assumed that the model is no longer learning. This minimizes (optimizes) computational resources and prevents over-fitting, which reduces the model's capacity to generalize.

A subclass of deep neural networks called convolutional neural networks (CNNs) was developed especially for tasks involving visual data, like object detection and image classification, among others. CNNs are robust deep learning models designed to handle visual data. They employ loss functions to regulate the training process, activation functions to introduce non-linearity, fully connected layers for classification, pooling layers for spatial reduction, and convolutional layers for feature extraction. The model's performance and generalization are enhanced by regularization methods including dropout and batch normalization.

3.6. System Implementation

Python was the programming language of choice for putting the optimized model into practice. The application is compatible with Linux operating systems that can run Debian files, such as Ubuntu and Debian, and Windows operating systems, ranging from Windows 7 to Windows 10. Since the application is stand-alone, neither mobile devices nor the Web are intended for it. The application can be used on a desktop computer or a laptop. A 1.3 GHz processor and 512 MB of Random Access Memory (RAM) are sufficient for the efficient operation of a 20–50 GB hard drive.

4. Results and Discussion

F-measure, precision, and recall are among the performance indicators of the models, while confusion matrices are used to show the outcomes. Equation 1 represents precision, which is defined as the ratio of correctly predicted positive observations to all expected positive observations. A low false positive rate is linked to good accuracy.

$$\text{Precision} = True\ Positive/(True\ Positive + False\ Positive) \tag{1}$$

Recall, also referred to as sensitivity, is defined as the ratio of correctly predicted positive observations to all of the actual class observations. It is as Equation 2 indicates.

$$\text{Recall} = True\ Positive/(True\ Positive + False\ Negative) \tag{2}$$

The F-measure is the precision and recall weighted average. Equation 3 illustrates it, and it is expressed in percentage terms.

$$\text{F} - \text{Measure} = 2 * \frac{(Recall * Precision)}{(Recall + Precision)} \tag{3}$$

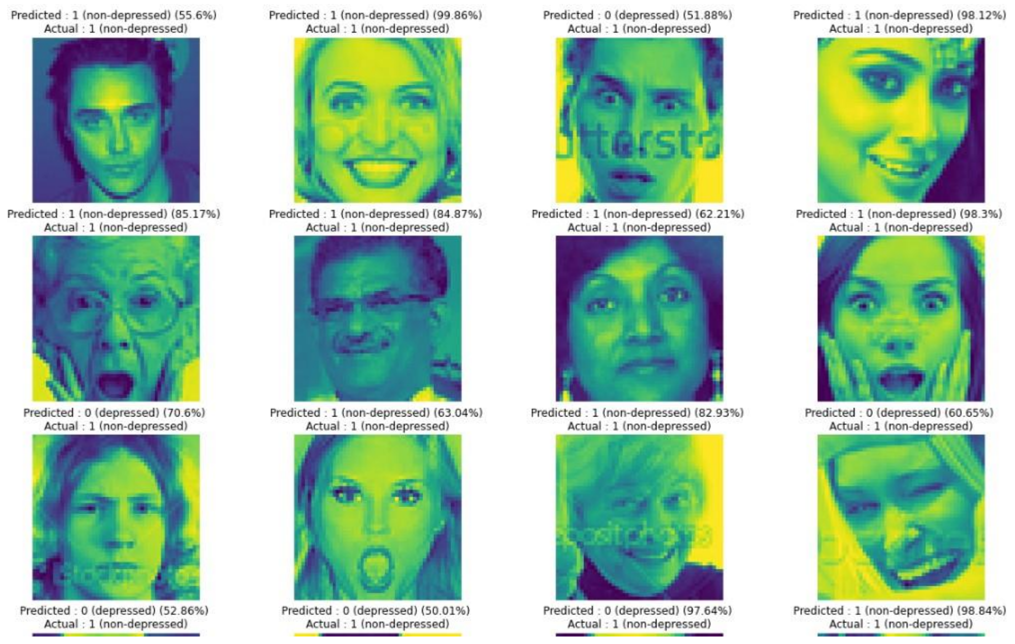


Fig.3. CNN prediction for non-depressive images

Fig.4. CNN prediction for depressive images

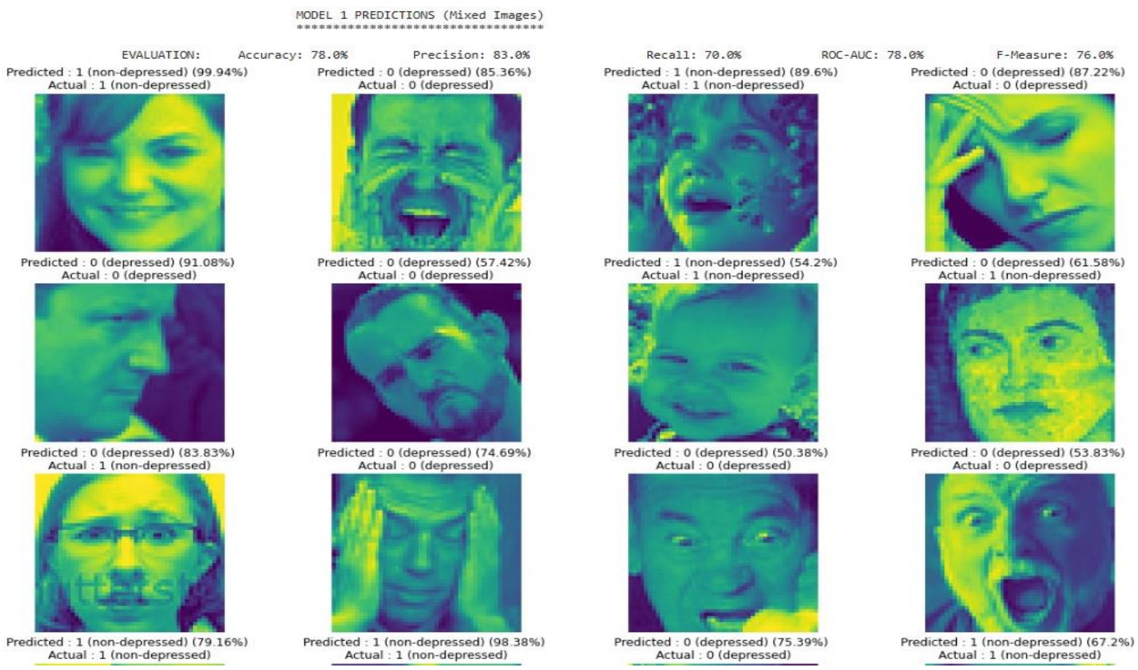


Fig.5. CNN prediction for mixed depressive and non-depressive images

The CNN prediction results for non-depressed images are displayed in Fig.3, while the CNN prediction results for depressive images are displayed in Fig.4. The prediction for mixed depressed and non-depressed images is displayed in Fig.5.

The results of the model are as shown in the Confusion Matrix on Fig. 6. It shows that 1,617 cases were correctly detected as having depressed symptoms and 1608 as having non-depressed symptoms. It also shows that 388 depressed cases were wrongly detected as non-depressed and 696 non-depressed were detected as depressed cases.

The evaluation results for the CNN model are shown in Table 5 and Fig. 7. The table shows that the model has an accuracy of 0.77, Precision of 0.81, Recall of 0.72, ROC-AUC of 0.77 and F-Measure of 0.76.

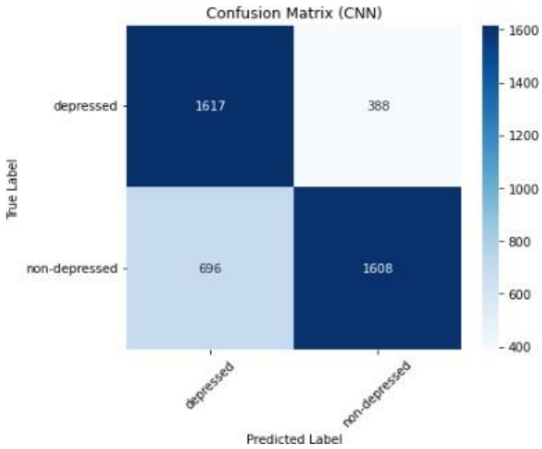


Fig.6. Confusion matrix for the CNN optimized model

Table 5. Evaluation results for CNN model

Metric	Result
Accuracy	0.77
Precision	0.81
Recall	0.72
ROC-AUC	0.77
F-Measure	0.76

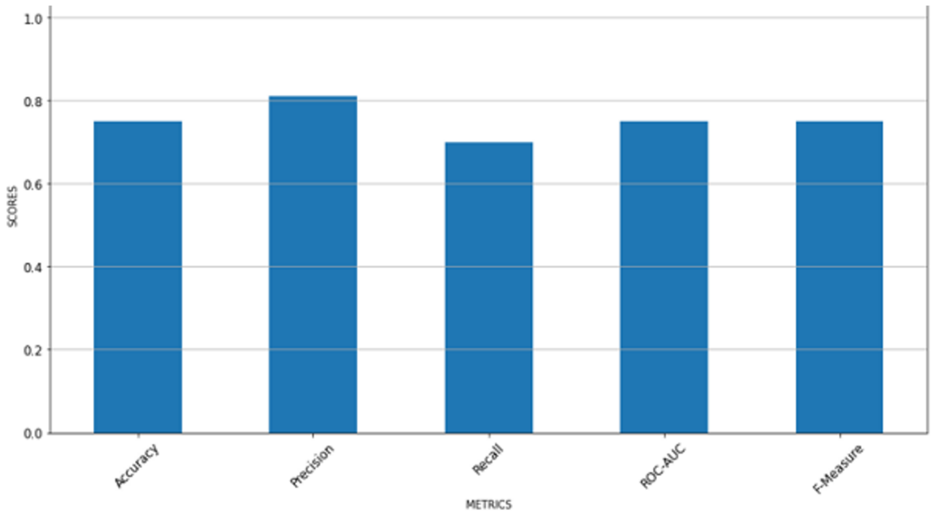


Fig.7. Evaluation results for the CNN model

The model achieves an overall accuracy of 0.77. This suggests that it correctly predicts the class for 77% of the instances.

The precision for the combined predictions is 0.81, indicating a good proportion of correct positive predictions.

The Recall of 0.72 means that the model captures 72% of the actual positive instances. The ROC-AUC value of 0.77 represents the model's ability to distinguish between the two classes. The F-measure of 0.76 provides a balanced measure of precision and recall.

These results suggest that the CNN model performs reasonably well in classifying depressive and non-depressive images with relatively balanced precision and recall values for both classes.

5. Conclusions

In this paper, an innovative methodology to detect depressive symptoms using Twitter dataset is presented. The study has proven that it can efficiently distinguish between depressive and non-depressive symptoms in image posts. It provides an optimized model that can predict distinct depressive groups from user-generated data. The results obtained show that the optimized model achieves an accuracy of 77%. This implied that the system achieves high accuracy. This study contributes to the body of knowledge by providing a methodology for detecting depressive symptoms in image posts. The system has the potential to be used as a tool for detecting depressive symptoms in social media posts, which could be useful for mental health professionals and researchers.

References

[1] Jeanine M. Mingé and Nicole. Defenbaugh, "Communicating and Navigating Digitized Healthcare". Storied Health and Illness: Communicating Personal, Cultural, and Political Complexities, pp 105, 2016.
[2] Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria P. Reyes, Mei-Ling Shyu, Shu. C. Chen, and Sundaraja S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications". ACM Computing Surveys (CSUR), 51(5):1–36, 2018.
[3] Chao Zhang, Zichao. Yang, Xiaodong He and Li Deng, "Multimodal intelligence: Representation learning, information fusion, and applications". IEEE Journal of Selected Topics in Signal Processing, 14(3):478–493, 2020.
[4] Mann Paulo, Aline. Paes and Elton H. Matsushima, "See and read: detecting depression symptoms in higher education students using multimodal social media data". In Proceedings of the International AAAI Conference on Web and Social Media, volume 14, pp. 440–451, 2020.
[5] James Lando, Sheree. M. Williams, Stephanie Sturgis and Branalyn Williams, "A logic model for the integration of mental health into chronic disease prevention and health promotion". *Preventing Chronic Disease*, 3(2), 2006.
[6] Michelle Morales, Stefan Scherer and Rivka Levitan, "A cross-modal review of indicators for depression detection systems". In Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality, pp. 1–12, 2017.
[7] Munmum De Choudhury, Michael Gamon, Scott Counts and Eric Horvitz, "Predicting depression via social media". In *Seventh International AAAI conference on weblogs and social media*, 2013.
[8] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh and Hiroyuki. Ohsaki (2015). Recognizing depression from twitter activity. *In Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp. 3187–3196.
[9] Zunaira Jamil (2017). Monitoring tweets for depression to detect at-risk users. PhD thesis, Universit´e d'Ottawa/University of Ottawa.
[10] Yu-Ching Huang, Chieh. F. Chiang and Arbee L. Chen (2019). Predicting Depression Tendency based on Image, Text and Behavior Data from Instagram. In *DATA*, pp. 32–40.
[11] Amir H. Yazdavar, Mohammad S. Mahdavinejad, Goonmeet. Bajaj, William Romine, Amit Sheth, Amir H. Monadjemi, Krishnaprassad Thirunarayan, J. M. Meddar, A. Myers, J. Pathak, "Multimodal mental health analysis in social media". Plos one, 15(4): e0226248, 2021.
[12] Bhumika Gupta, Neeraj Pokhriyal and Kamal K. Gola, "Detecting Depression in Reddit Posts using Hybrid Deep Learning Model LSTM-CNN". In *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, pp. 610–617, IEEE. 2022.
[13] Harnain Kour and Manoj K. Gupta, "A hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM". *Multimedia Tools and Applications*, 81(17):23649–23685, 2022.
[14] Tolulope M. Awoyelu, Abimbola. R. Iyanda and Sunday K. Mosaku, "Formulation of a Predictive Model for the Determination of Depression among University Students". In Proceeding of the 14th International Conference on Smart Nations, Digital Economies and Meaningful Lives Nigerian Computer Society, (NCS.2019), pp. 098– 109, 2019.
[15] Panch Ratan, "What is Convolutional Neural Network Architecture?". *Analytics Vidhya*, 2020. Available at: https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/.Accessed: May 27th 2023.

Authors' Profiles

Awoyelu T. M. is an Assistant Lecturer in the Department of Computer Sciences, Faculty of Natural Sciences, Ajayi Crowther University, Oyo Township, Oyo State. She obtained her B.Sc. (Hons) Degree of Computer Science from the Department of Computer Science and Engineering, Osun State University, Osogbo in 2014. She also holds M.Sc. degree in Intelligent Systems Engineering from the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife in 2018. She has over 5 years of teaching and research experience. Her research interests are in Artificial Intelligence and Data Analytics. She has successfully supervised 10 Undergraduate Theses. She has five publications to her credit. She is a member of Nigeria Computer Society (NCS) and Nigerian Women in Information Technology (NIWIIT).

Dr. Abimbola R. Iyanda holds a B.Sc. degree in Computer Engineering, M.Sc. and Ph. D. degrees in Computer Science from Obafemi Awolowo University, Ile-Ife, Nigeria. The thrust of her research is in the area of Computing and Intelligent Systems Engineering with a focus on Natural Language Processing and Human Computer Interactions aiming at domesticating computer technology and the computational rendering of indigenous ideas. She is a Member of the Nigerian Society of Engineers, Association of Professional Women Engineer in Nigeria, Council for the Regulation of Engineering in Nigeria, Organization for Women in Science for the Developing World (OWSD) and Nigeria Computer Society. Her present employment is with the Computer Science and

Engineering Department, Obafemi Awolowo University, Ile-Ife, Nigeria.

Prof. S. K. Mosaku is a professor of Mental Health. He is a member of staff of the Department of Mental Health, Obafemi Awolowo University, Ile-Ife, Osun state, Nigeria. His international experience includes various programs, contributions and participation in different countries for diverse fields of study. His research interests reflect in his wide range of publications in various national and international journals. He is the Editorial Board Member of Journal of Psychiatry. His research interests are Posttraumatic stress, Anxiety and Depression.

How to cite this paper: Awoyelu T. M., Iyanda A. R., Mosaku S. K., "An Optimized Convolutional Neural Network Model for Detecting Depressive Symptoms from Image Posts", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.4, pp.45-55, 2024. DOI:10.5815/ijitcs.2024.04.03