HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Detecting and Validating the Emotions in Alzheimer's Patients by Voice Analysis, Computer Vision and Deep Learning

Krishnni Khanna[1] and Reetu Jain[2]

[1]Westwood High School, USA
[2]On My Own Technology Pvt Ltd, India

ABSTRACT

Alzheimer's disease is a degenerative disorder of the brain that affects memory and cognitive function, and is becoming more prevalent as the population ages. Alzheimer causes brain cells of a person to die, and with time the brain works less. As a result of this there is a change in the behavior and personality of Alzheimer patients. It is observed that the patients very often suffer from fluctuating mood. Due to the change in mood of the patient, the caretaker or the attendant of the patient is unable to distinguish what triggered the changes, and provide proactive support or timely intervention. Hence the study is undertaken to detect the mood of the Alzheimer's patient employing the voice analysis, computer vision and deep learning. This paper comprises of two phases. The first phase detects the emotions of the patients in real time with the help of computer vision (CV), voice analysis (VA) and Convolutional Neural Network (CNN). Two CNN models were trained, first with the attributes extracted from the image and second the features extracted from the voice dataset. Then the predictions are compared to get the result from the proposed model. The second phase of the paper examines the practicality of the proposed approach by applying it to detect the emotions in four Alzheimer's patients. Finally, the results are compared and validated. Overall, the proposed model holds promise as a valuable tool in the real time detection of emotions in Alzheimer patient, enabling timely intervention and improved patient outcomes.

Introduction

Alzheimer's disease is a neurodegenerative disorder characterized by progressive cognitive decline and memory loss, along with other symptoms such as language difficulties, disorientation, and mood and behavior changes [1]. It is the most common cause of dementia, and is believed to be caused by a combination of genetic, environmental, and lifestyle factors. Anxiety and depression are common symptoms in Alzheimer patients, often due to the loss of cognitive and functional abilities and the stress of coping with the disease [2]. These symptoms can exacerbate other cognitive and behavioral symptoms of Alzheimer's and impact the overall quality of life for patients. It's important to monitor and address these symptoms with a combination of medication, therapy, and lifestyle changes to improve patients' well-being.

Sudden mood swings and personality changes are very often observed in the Alzheimer's patients. As a result of which the caretaker or the attendant of the patient is unable to provide proper support to the patients. Many scientists and researchers are researching to identify reasons that trigger the sudden mood swings in the Alzheimer's patients. These researches are not only helpful in detecting anxiety in patients but can also improve their quality of life, prevent complications, and improve caregiver support [3].

Motivation and Novelties

Motivated by the researches in detecting the emotional and personality changes in the Alzheimer's patients, this study employs convolutional neural networks (CNN), computer vision (CV) and voice analysis in detecting emotions of patients in real time. From the literature reviewed in the domain of present studies, it is observed that there are very few researches that employ machine learning (ML) or deep learning (DL) algorithms for analysis. However, the papers that do employ ML or DL algorithms, they either have low accuracy or take high computation time. To address these concerns, our study in this paper integrates CV and voice analysis with CNN for real time detection of the emotions and personality in the Alzheimer's patients. The analysis can be performed with better accuracy and lesser computation time compared to previous similar studies. The strength of the proposed approach is its practical applicability. Hence, it is applied to ten test cases that includes Alzheimer's patients as well as person without Alzheimer.

Hereon, this paper is structured as follows:

Section 2 presents the review of the contemporary literatures. Section 3 describes the preliminary concepts required to develop the proposed methodology in section 3. Section 4 presents the results and discussions of our study, and finally section 5 presents conclusion.

Review of The Contemporary Literatures

Speech analysis can be used to detect Alzheimer's disease by analyzing various features of speech of the patient, such as the use of vocabulary, grammar, and syntax. One common feature of Alzheimer's disease is a decline in ability to use language, including a reduction in vocabulary size, difficulty with word retrieval, and decreased syntactic complexity [4]. ML algorithms can be trained on a dataset of speech samples, with some labeled as having Alzheimer's and others labeled as not having Alzheimer's, to learn these patterns and to then predict whether a new speech sample indicates Alzheimer's.

Several studies have shown that ML algorithms can accurately detect Alzheimer's disease based on speech samples [5]. For example, a study published in the Journal of Alzheimer's Disease in 2016 used an ML algorithm to analyze speech samples from participants with and without Alzheimer's disease. The algorithm was able to accurately identify participants with Alzheimer's disease with a sensitivity of 84.6% and a specificity of 90.3% [6]. Another study published in 2021 used an ML algorithm to analyze speech samples from participants with and without Alzheimer's disease. The algorithm was able to accurately identify participants with Alzheimer's disease with a sensitivity of 80% and a specificity of 96% [7]. Based on these published studies, it can be concluded that speech analysis has the potential to be a valuable tool in detecting Alzheimer's disease. However, it is important to note that speech analysis cannot substitute professional diagnosis and treatment.

ML techniques can be used to detect anxiety by analyzing various data inputs such as speech, text, or physiological signals. One example of using ML to detect anxiety is through analyzing speech patterns. Research has shown that people with anxiety tend to speak in higher pitch, they speak more slowly, and use more fillers such as "uh" and "um” [8]. ML algorithms can be trained on a dataset of speech samples, with some labeled as having anxiety and others labeled as not having anxiety, to learn these patterns and predict whether a new speech sample indicates anxiety.

Another example is analyzing text data, such as social media posts or emails. People with anxiety tend to use more negative words and express more worry or fear in their language. ML algorithms can be trained on a dataset of text samples, labeled with anxiety or non-anxiety, to learn these patterns and make predictions about new text data [9]. Physiological signals can also be analyzed to detect anxiety, such as heart rate variability and skin conductance. ML algorithms can be trained on a dataset of physiological signals, labeled with anxiety or non-anxiety, to learn the patterns associated with anxiety and predict the presence of anxiety in new signals.

Speech recognition is one of the most common applications of ML in speech analysis. ML algorithms can be trained on a large dataset of speech samples and their corresponding transcriptions, allowing the algorithm to learn the patterns and relationships between spoken words and their corresponding text. This can enable accurate speech-to-text conversion, which is useful in applications such as virtual assistants, automated transcription, and closed captioning [10].

Emotion recognition is another application of ML in speech analysis. ML algorithms can be trained on a dataset of speech samples labeled with emotional states, such as happy, sad, angry, or neutral. The algorithm can learn the patterns of speech that correspond with each emotional state, allowing it to accurately classify new speech samples based on the emotional content [11].

Speech synthesis is a yet another application of ML in speech analysis. ML algorithms can be trained on a dataset of speech samples and their corresponding text transcriptions, allowing the algorithm to learn the patterns and relationships between text and speech. This can enable the algorithm to generate natural-sounding speech from text, which is useful in applications such as text-to-speech systems and automated voice assistants [12].

Material and Methods

In this section of the paper, preliminary concepts employed for the study, the dataset used for training the DL methods, and the proposed methodology for the study is discussed.

Dataset

In this study, both audio and image datasets were used for detecting the emotions in Alzheimer patients. Two different audio datasets were used for the process. The first dataset is the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [13]. This dataset comprised of the 1500 audio files with both male and female voices labeling 8 different emotions namely neutral, calm, happy, sad, angry, fearful, disgusted and surprised. The second audio dataset is the Surrey Audio-Visual Expressed Emotion (SAVEE) Database [14] that contains 500 audio files recorded in 4 different male voices with the above-mentioned emotions. The image dataset is the FER2013 [15] collected from Kaggle database [16]. The dataset comprises of 32,298 grayscale images of faces, each made of 48*48 pixels, which are classified into seven categories (angry, disgusted, fear, happy, sad, surprised, and neutral).

Preprocessing of the Data

The audio and image dataset used for developing the proposed approach contains some irregularities such as the number of labels in the RAVDESS and the FER2013 dataset comprises of different number of labels. Since 'neutral' and 'calm' labelled audio files did not show much variation, both audio file sets have been merged and labelled as 'neutral'. Additionally, since Alzheimer's predominantly appears in persons above 55 years of age, the images of persons less than 50 years have been dropped from image dataset. The resulting dataset comprised of 15,218 images. Out of which 13,427 images are used for training and the remaining 1791 images are used for testing.

Convolutional Neural Network (CNN)

CNN belongs to the family of neural networks. Unlike other neural networks, CNN is mostly used for classifying images based on the different parameters on which the images are trained on. The most important feature of CNN is that it is programmed to automatically and adaptively learn the order of the attributes of the input data, which makes them very effective in recognizing patterns in the input data.

A typical CNN model comprises of three layers namely convolution, pooling and fully connected layer [17]. The importance of convolution layer is to extract the attributes from the input data and to activate it using the rectified linear unit (ReLU) activation function. The ReLU activates the convoluted matrix by nullifying the noise in the input data [18]. The activated matrix is passed through the pooling layer to reduce the number of features' map dimensions. The max pooling operation is performed in this study. The convolution, activation and pooling operations are repeated until a better accuracy is obtained by the CNN model. This process is called layer stacking [19]. In order to increase the accuracy of the CNN model and detection in the real time, the convolution, activation and pooling operations are repeated for 3 times. After this the series of different attributes are linked to the neurons in the fully connected layer. The neurons represent a label for the input data. The label of the neuron that is activated by the series of attributes of the input data is the classification done by the CNN model [20]. Figure 1 shows a pictorial representation of the developed CNN model.

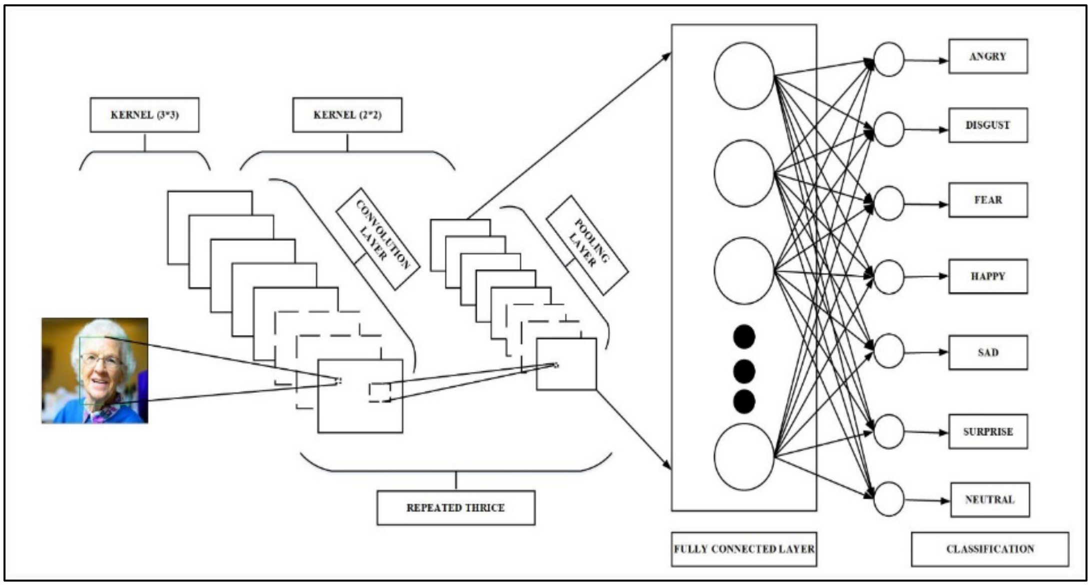

Figure 1. Pictorial representation of the CNN model

Voice Analysis (VA)

VA is the process of using technology to analyze and interpret various characteristics of human speech. This can include a wide range of tasks, such as speech recognition, speaker identification, emotion detection, and language translation [21]. Emotion recognition is perhaps the most well-known application of voice analysis. In this study, emotion detection is done by VA in integration with CNN.

The basic idea behind using a CNN for VA is to transform the audio signal into a spectrogram or other time-frequency representation that can be treated as a 2D image. The CNN is then trained to extract features from this input that are relevant to the emotion detection.

The Proposed Methodology

The proposed methodology involves the application of CV along with CNN for real time detection of emotions which is further validated by VA integrated CNN model. The steps for the proposed model is as follows:

Step 1: Loading of the data: In this step, the audio data are loaded using the librosa python library. On the other hand, the image dataset are loaded with the help of dataloader function of the pytorch library.

Step 2: Extracting features: In this step, the features from the audio file are extracted again using the librosa library in form of mel-frequency cepstral coefficients (MFCCs). The audio files split into fragments

HIGH SCHOOL EDITION
Journal of Student Research

with width of 25ms and a total of 39 features are extracted from each fragment. Splitting the audio into fragments increases the noise at the edges which were reduced by employing Hamming/Hanning windows as it also prevented the noise in the high-frequency region. Hamming and Hanning windows function are the most commonly used for random signals due to its ability to provide a good frequency resolution and leakage protection with fair amplitude accuracy [23]. The data extracted from the audio files are stored in a comma separated value (csv) file.

Step 3: Splitting the dataset: The FER2013 dataset are already split into training and testing dataset. As a result, no further split is done. On the other hand, the features extracted from the audio files are randomly divided in the ratio of 70:15:15. The splitting is done based on the higher testing accuracy. Where 70% of the data are used for training, 15% are used for testing and the remaining 15% of the data are used for validating the CNN model.

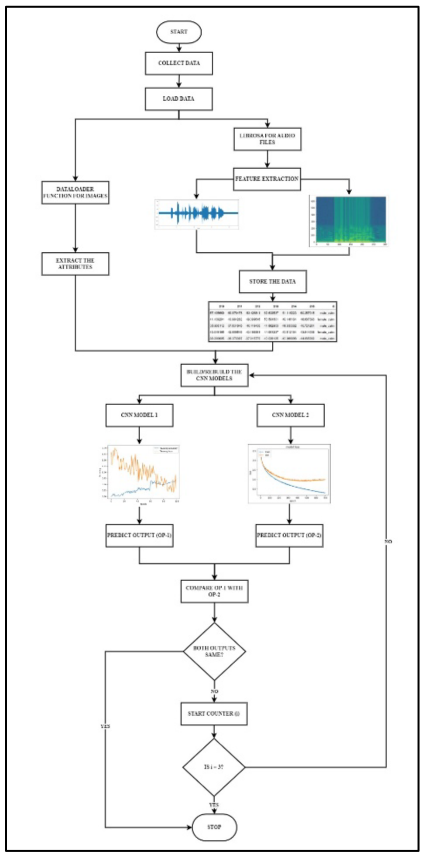

Figure 2. Flowchart of the proposed approach

Step 4: Training the dataset: Two separate CNN models are developed for training the image and audio datasets. The image dataset were trained for 100 epochs with 15 steps per epoch whereas the audio features dataset were trained for 100 epochs.

Step 5: Testing the datasets: The two CNN models created in step 4 are tested using the test dataset.

Step 6: Validating the results: In this step the result obtained from both the CNN models for a practical case is compared. In case the result from both the CNN models are same then the result is the label that is predicted. In case the result is different than the models will try to predict for two more times at an interval of 5 second each. If no common prediction is achieved from both the models after the third try then the output obtained will be “Could not identify”.

The flowchart for the proposed approach is shown in figure (2).

Results and Discussions

This section of the paper notes down the results obtained from the applying the proposed approach in detecting the emotions in the Alzheimer patients. The coding for the proposed approach is done using the python 3.10 in a Windows 11 PC with i5 processor and 8GB RAM.

Results

The Architectural Details of the CNN Model

Architecture of CNN model comprises of different parameters such as number of layer stacking, batch size, optimizer, learning rate etc. which are to be optimized before been applied to achieve the aim of the study. Moreover, CNN architectures vary in terms of depth and complexity. The deeper or wider the CNN model developed, the complexity of the model increases. Hence, in order to get a favorable result at a fair time period a smaller network is developed that returns preferably a good result. The wider network will take longer time to train. Deep networks are very computationally expensive to train. Hence, the CNN model developed in the study is made wide and deep enough that they return a good result. The different parameters in the study are fixed by hit and trial methods. By varying all the different parameters of CNN models and tabulating the training and testing accuracy as well as the CPU timing, the optimized architecture of the CNN model is determined. The optimized CNN parameters for audio and image files are shown in table 1 and 2 respectively.

Table 1. Optimized CNN parameters for Audio files training

Parameters	Implemented Layers / Stage	Values
Dimensions	Input layer	3
Input size	Input layer	12
Number of convolution kernels	Convolutional Layer	256
Convolution kernel size	Convolutional Layer	3
Pooling window size	Pooling Layer	2
Number of dense units	Fully-Connected Layer	7
Training:Testing:Validation	Data preparation	70:15:15
Learning rate	Learning phase	0.015

Table 2. Optimized CNN parameters for image files training

Parameters	Implemented Layers / Stage	Values
Dimensions	Input layer	2
Input size	Input layer	9
Number of convolution kernels	Convolutional Layer	144
Convolution kernel size	Convolutional Layer	4
Pooling window size	Pooling Layer	3
Number of dense units	Fully-Connected Layer	7
Training:Testing:Validation	Data preparation	70:15:15
Learning rate	Learning phase	0.005

Results Obtained from the CNN Models

The first step of the proposed methodology is to load the files followed by extracting the features of the audio files and to plot the waveform and spectrogram. Figure (3) and (4) shows the plotting of an input audio in the waveform and spectrogram respectively for a first fragment of a happy audio file. The features extracted are stored in a csv file which are mapped with the different labels namely angry, disgusted, fear, happy, sad, surprised, and neutral.

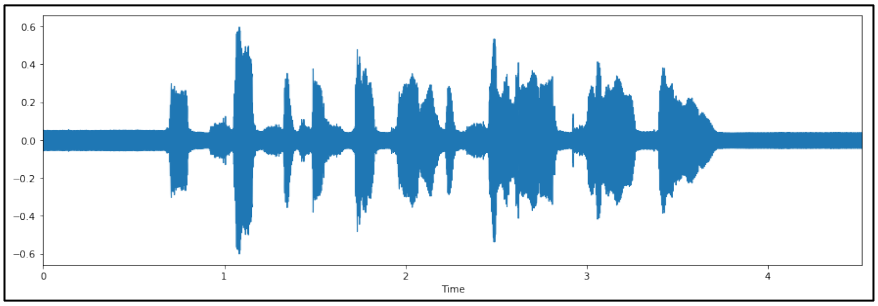

Figure 3. Waveform plot for an audio input

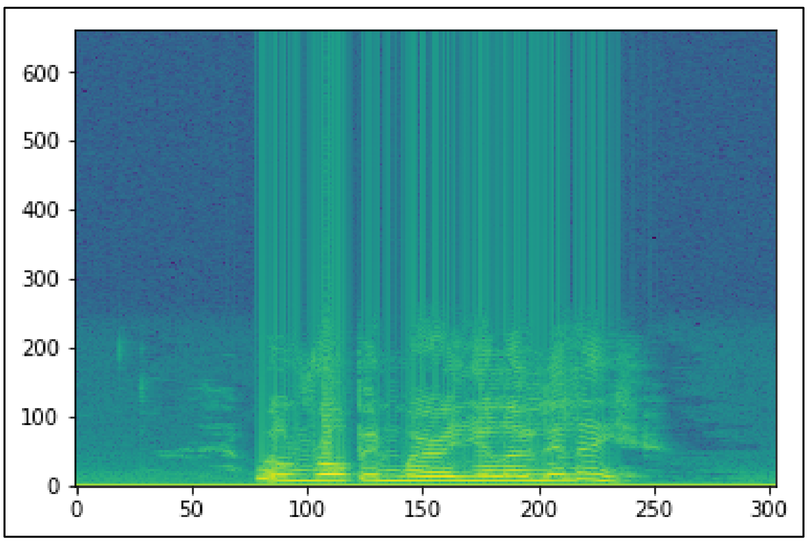

Figure 4. Spectrogram plot for the audio input

The third step of the proposed approach is to split the data accordingly as mentioned in section 3.4 of the paper. In the fourth step, the CNN models are trained for the image and audio datasets. The image dataset are trained for 100 epochs with 15 steps per epochs whereas audio dataset are trained for 700 epochs. The convolution, activation and pooling is repeated thrice and four times for the image and audio dataset respectively. For both the datasets, ReLU activation function is employed for transferring the convolution matrix whereas Softmax activation function is employed for computing the probability of occurring of the different labels. The weights of the CNN models are optimized by the Adaptive Moment Estimation or the Adam optimizer [22]. The step 5 of the proposed approach is to test the CNN models developed for the study. Figure (5) and (6) shows the training and testing accuracy and model loss curve for the image files.

HIGH SCHOOL EDITION
Journal of Student Research Volume 13 Issue 1 (2024)

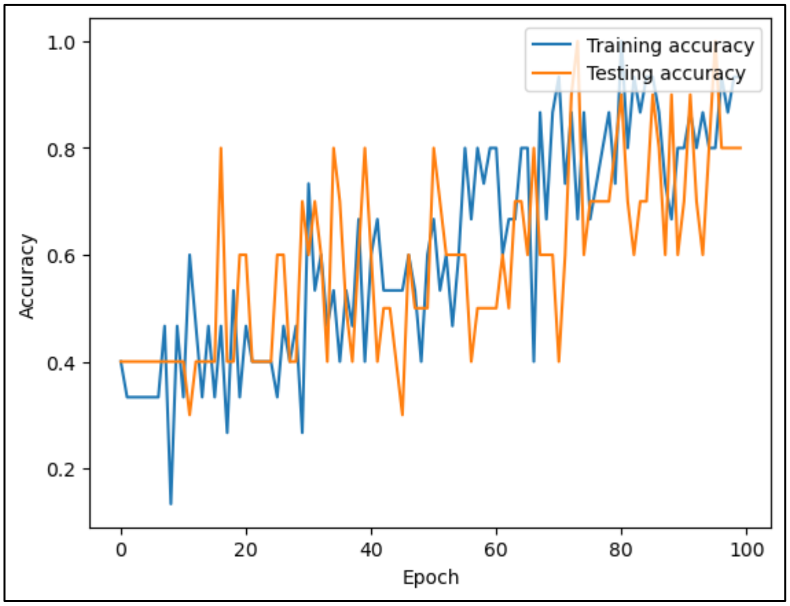

Figure 5. Training and testing accuracy curve for the FER2013 dataset

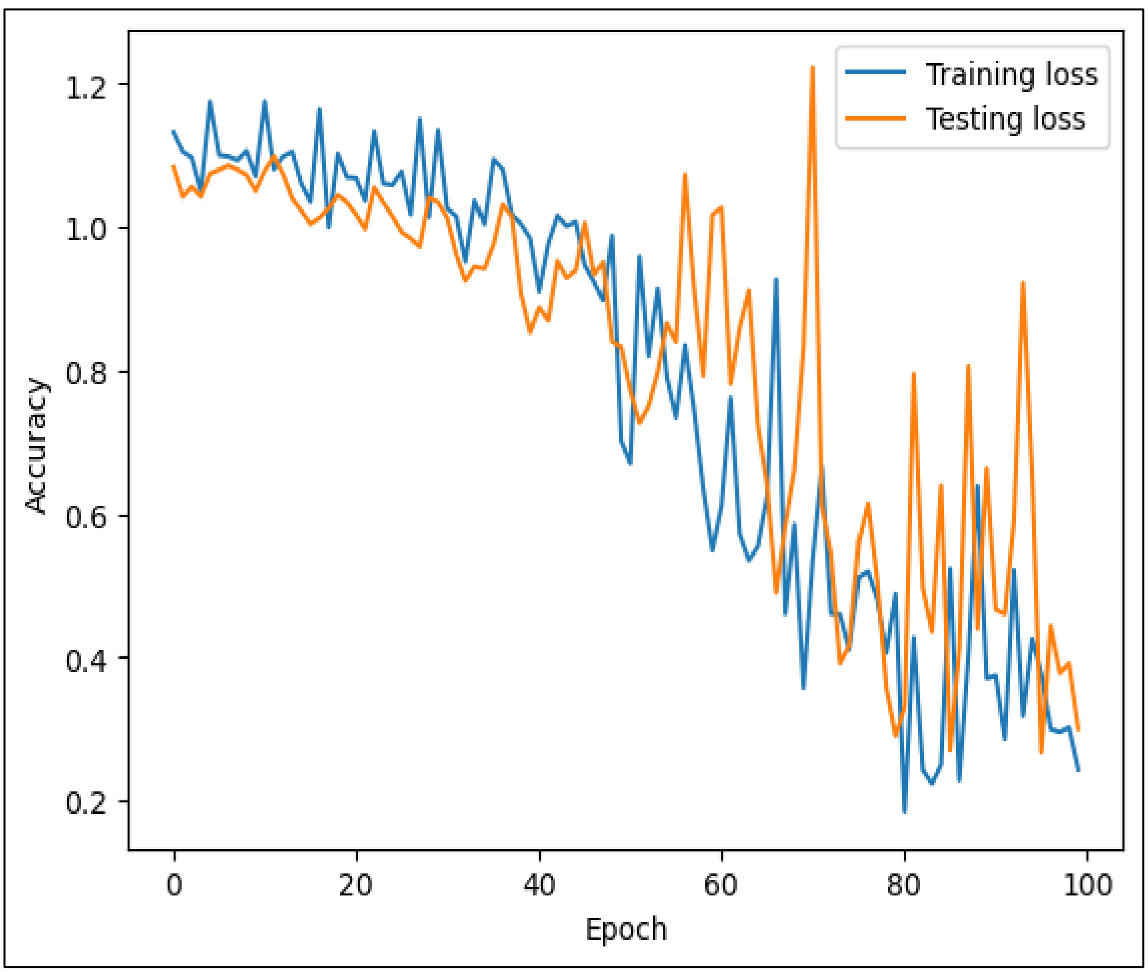

Figure 6. Testing and training loss curve for the FER2013 dataset

It can be observed from the figure (5) and (6) that there is a huge fluctuation in the graphs of the training and testing loss. This is due the fact that the FER2013 dataset is large and the size of the images is very small. The training and the testing accuracy for the first iteration is 0 which increased to 0.66 and 0.61 at the 100th iteration. The training and testing loss for the CNN model is 1.807 and 2.86 in the 1st iteration which decreased to 0.32 and 0.79 in the 100th iteration. The testing dataset comprises of 1791 images where all the labels comprises of 256 images except for the neutral label with 255 images. For validating the CNN model, the confusion matrix is plotted and shown in figure (7). The developed CNN model correctly identified 58%, 63%, 52%, 85%, 59%, 73% and 62% of the angry, disgusted, fear, happy, sad, surprised and neutral labels respectively. Whereas the model misclassified the remaining.

The figure (8) shows the training and testing model loss for the CNN model created for the audio dataset. The CNN model is trained for 100 epochs. The training and testing loss values for the model was 1.4 and 1.2 which lowered to 0.6 and 0.65 respectively at the 100th epoch. For validating the CNN model, the confusion matrix is plotted and shown in figure (9). The developed CNN model correctly identified 55%, 62%, 57%, 72%, 81%, 63% and 69% of the angry, disgust, fear, happy, sad, surprise and neutral labels respectively. Whereas the model misclassified the remaining.

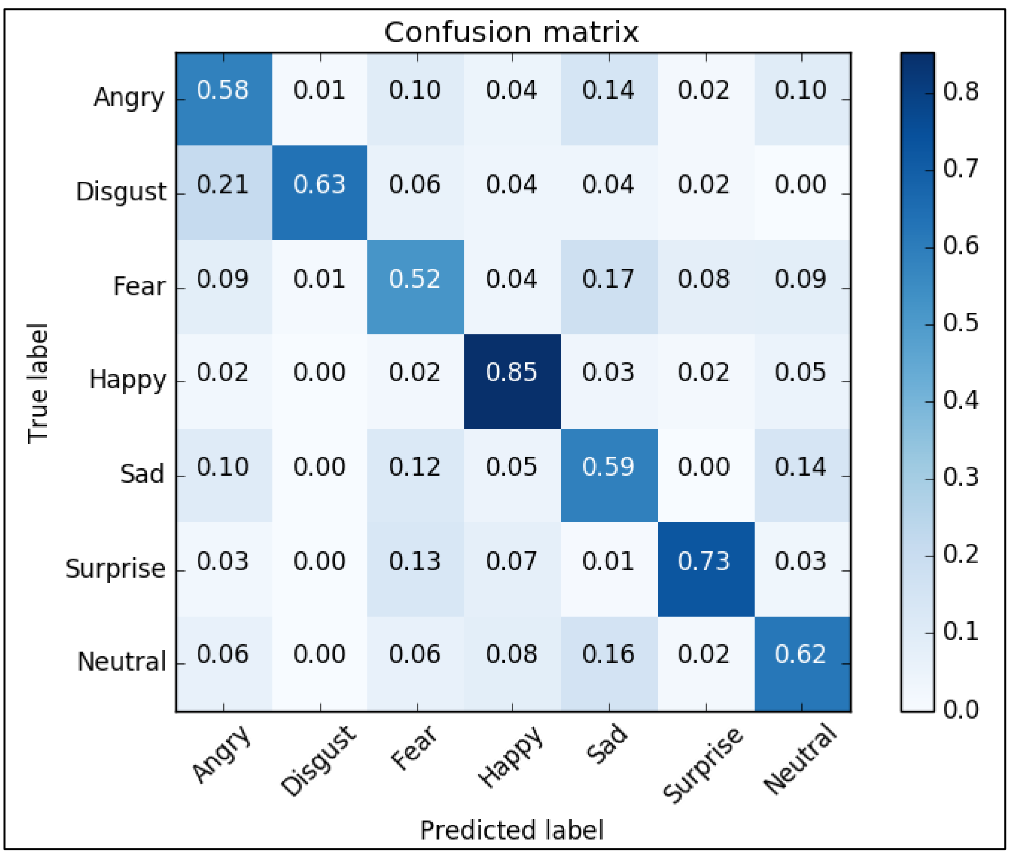

Figure 7. Confusion matrix of the FER2013 dataset

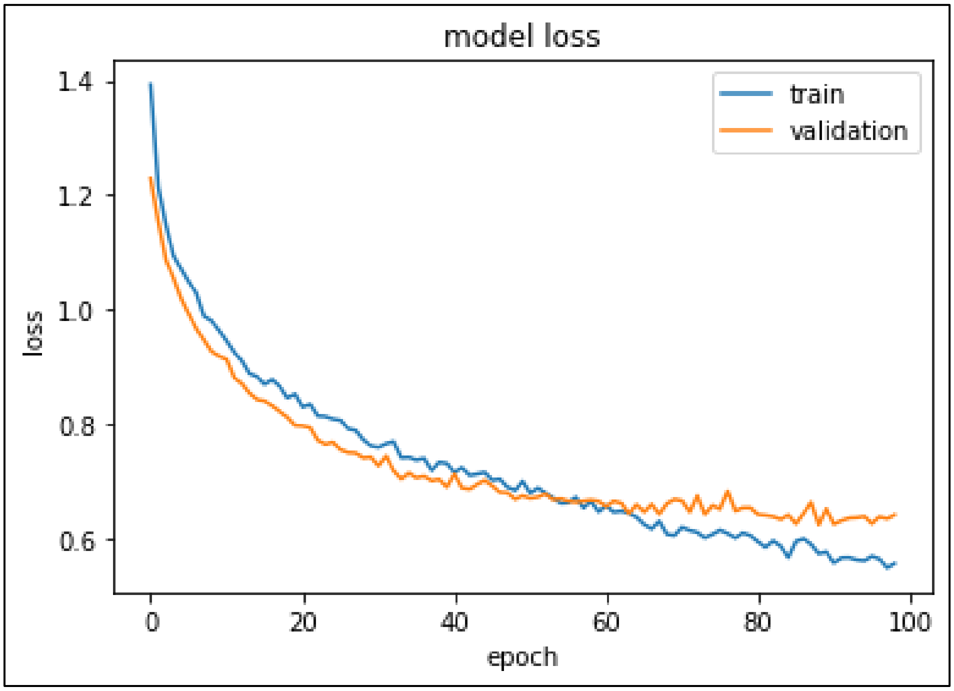

Figure 8. Training and testing loss value for the CNN model

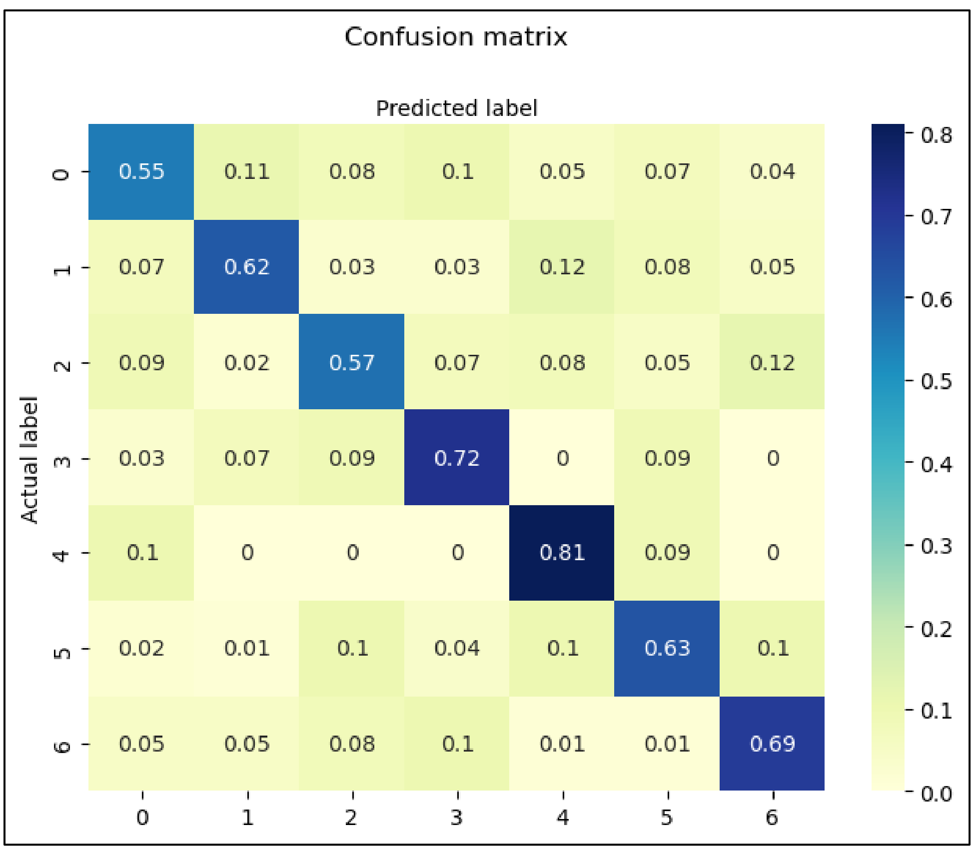

Figure 9. Confusion matrix of the audio dataset

In figure (9), the emotions angry, disgusted, fear, happy, sad, surprised and neutral labels are indexed from 0 to 6 respectively. The final step of the proposed approach is to validate the results. In this step the developed CNN models are tested for detecting the emotions in Alzheimer's patients. If both the CNN models, predicted the same label then that label is the output received from the proposed approach. If the predicted labels are different in both the CNN models then the prediction process is repeated three more times. Then, if the same label is not obtained by the CNN models, the proposed approach exits with "Could not identify".

Testing and Application of the Proposed Approach

To test the feasibility of the proposed model, it is tested in 10 different instances to predict the emotion of four Alzheimer patient named as AP-1, AP-2, AP-3 and AP-4. The result obtained for all the 10 instances are summarized in table 3.

Table 3. Summary table

Sl. No.	Patient	Output from		Result	Actual emotion	Hit/Miss
		CNN model 1	CNN model 2			
1	AP-1	Happy	Happy	Happy	Happy	Hit
2	AP-1	Happy	Sad	Could not identify	Neutral	Miss
3	AP-2	Fear	Fear	Fear	Sad	Miss
4	AP-3	Angry	Disgust	Could not identify	Disgust	Miss
5	AP-3	Angry	Angry	Angry	Angry	Hit
6	AP-3	Sad	Sad	Sad	Fear	Miss
7	AP-4	Happy	Happy	Happy	Happy	Hit
8	AP-4	Neutral	Neutral	Neutral	Neutral	Hit
9	AP-4	Surprise	Surprise	Surprise	Surprise	Hit
10	AP-4	Sad	Sad	Sad	Sad	Hit

CNN model 1 and CNN model 2 are models developed by training the FER2013 and audio dataset respectively. From table 1, it is observed that out of the 10 instances, the proposed approach was able to correctly identifying the correct emotions for 6 times. It is also observed that the result from proposed approach could not identify for two instances. Moreover, it can also be observed that the CNN model-1 correctly identified 6 times whereas CNN model-2 was correctly identified 7 times. The CNN model-2 showed more accuracy than the CNN model-1. From the analysis, it can be ascertained that the proposed model can be used for detecting the emotions in Alzheimer patients in real time.

Conclusions

Alzheimer is a neurodegenerative disorder often characterized by sudden change in emotions, behavior, and memory loss and so on. Researchers are constantly working on finding new ways of identifying the emotions in them. These researches will be helpful in improving the quality of life for the Alzheimer patient and will also be helpful for the caretaker in providing the requisite support. In this regard, the study uses CNN to detect the emotions. Two different CNN models were created where the first is trained using image dataset and the second is trained using an audio dataset. The output predicted by both of the CNN models are compared. In order to test the practicality of the proposed approach, it is tested for four Alzheimer patients for 10 instances. The proposed approach correctly identified 6 instances, whereas it missed for the remaining 4. Both the CNN models developed for the study showed some similar accuracy which is evident from the confusion matrix as shown in figure (7) and (8). In order to avoid any overfitting k-fold cross validation is done. The result obtained from the study showed a perfect-simple result which is returned within fairly a small time interval. Hence, from the overall discussions it can be concluded that the model could be used for detecting the emotions in Alzheimer's patients.

Author Contributions

K conceptualized the project, developed the proposed methodology, wrote the paper, and computed the solution. RJ supervised and guided the project. All authors reviewed and revised the manuscript.

Funding

On behalf of all authors, the corresponding author states that no funding in any form is received from any organization for carrying out the research work.

Data Availability

The data that support the findings of this study are enclosed as supplementary materials along with the manuscript.

Declarations

Conflict of interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Mecocci, Patrizia, and Maria Cristina Polidori. Antioxidant clinical trials in mild cognitive impairment and Alzheimer's disease. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1822.5 (2012): 631-638. https://doi.org/10.1016/j.bbadis.2011.10.006

Hort, J. O. B. J., O'brien, J. T., Gainotti, G., Pirttila, T., Popescu, B. O., Rektorová, I., ... & EFNS Scientist Panel on Dementia. (2010). EFNS guidelines for the diagnosis and management of Alzheimer's disease. European Journal of Neurology, 17(10), 1236-1248. https://doi.org/10.1111/j.1468-1331.2010.03040.x

Leung, D. K., Chan, W. C., Spector, A., & Wong, G. H. (2021). Prevalence of depression, anxiety, and apathy symptoms across dementia stages: A systematic review and meta-analysis. International Journal of Geriatric Psychiatry, 36(9), 1330-1344. DOI:10.3233/JAD-200739
Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. Journal of Alzheimer's Disease, 49(2), 407-422. DOI:10.3233/JAD-150520
Al-Hameed, S., Benaissa, M., & Christensen, H. (2016, September). Simple and robust audio-based detection of biomarkers for Alzheimer's disease. In 7th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT) (pp. 32-36). DOI:10.21437/SLPAT.2016-6
Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. Journal of Alzheimer's Disease, 49(2), 407-422. DOI:10.3233/JAD-150520
El-Sappagh, S., Alonso, J. M., Islam, S. M., Sultan, A. M., & Kwak, K. S. (2021). A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. Scientific reports, 11(1), 1-26. DOI:10.1038/s41598-021-82098-3
Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. Cognition, 84(1), 73-111.
Lee, T. R., Kim, G. H., & Choi, M. T. (2022). Identification of Geriatric Depression and Anxiety Using Activity Tracking Data and Minimal Geriatric Assessment Scales. Applied Sciences, 12(5), 2488 DOI:10.3390/app12052488
Li, J., Deng, L., Haeb-Umbach, R., & Gong, Y. (2015). Robust automatic speech recognition: a bridge to practical applications. DOI:10.1109/TASLP.2022.3153265
Shen, P., Changjun, Z., & Chen, X. (2011, August). Automatic speech emotion recognition using support vector machine. In Proceedings of 2011 international conference on electronic & mechanical engineering and information technology (Vol. 2, pp. 621-625). IEEE. DOI:10.1109/EMEIT.2011.6023178
Kim, J., Kim, S., Kong, J., & Yoon, S. (2020). Glow-tts: A generative flow for text-to-speech via monotonic alignment search. Advances in Neural Information Processing Systems, 33, 8067-8077. DOI:10.3390/app12052490
Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9, 611-629. DOI:10.1007/s13244-018-0639-9
Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PloS one, 13(5), e0196391. DOI:10.1371/journal.pone.0196391
Jackson, P., & Haq, S. (2014). Surrey audio-visual expressed emotion (savee) database. University of Surrey: Guildford, UK.FER-2013, https://www.kaggle.com/datasets/msambare/fer2013 Extracted on 20th February 2023. DOI:10.20944/preprints202309.1202.v1
Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9, 611-629. https://doi.org/10.1007/s13244-018-0639-9
Shyam, R. (2021). Convolutional neural network and its architectures. Journal of Computer Technology & Applications, 12(2), 6-14p. DOI:10.37591/JoCTA
Cen, L. P., Ji, J., Lin, J. W., Ju, S. T., Lin, H. J., Li, T. P., ... & Zhang, M. (2021). Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. Nature communications, 12(1), 4828. DOI:10.1038/s41467-021-25138-w
Dabiri, S., & Heaslip, K. (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. Transportation research part C: emerging technologies, 86, 360-371. https://doi.org/10.1016/j.trc.2017.11.021
Mishra, A., & Shukla, A. (2020, December). Psychological determinants of consumer's usage, satisfaction, and word-of-mouth recommendations toward smart voice assistants. In Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation: IFIP WG 8.6 International Conference

on Transfer and Diffusion of IT, TDIT 2020, Tiruchirappalli, India, December 18–19, 2020, Proceedings, Part I (pp. 274-283). Cham: Springer International Publishing. DOI:10.1007/978-3-030-64849-7_24

ŞEN, S. Y., & ÖZKURT, N. (2020, October). Convolutional neural network hyperparameter tuning with adam optimizer for ECG classification. In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-6). IEEE. DOI:10.1109/ASYU50717.2020.9259896

He, X. S., He, F., & Cai, W. H. (2016). Underdetermined BSS based on K-means and AP clustering. Circuits, Systems, and Signal Processing, 35, 2881-2913. DOI:10.1007/s00034-015-0173-7