International Journal of Data Science and Advanced Analytics (ISSN: 2563-4429)

Predicting Aggressive Behavior in Dementia Patients Using Text Classification with Word2Vec-LSTM

Shamaila Iram[1], Rejeesh Thayyil[2] and Hafiz Muhammad Athar Farid[3]
[1, 2, 3] Department of Computer Science, University of Huddersfield, United Kingdom.
[1] s.iram@hud.ac.uk, [2]rejeeshthayyil@gmail.com , [3]hmatharfarid@gmail.com

Corresponding author email: hmatharfarid@gmail.com

***Abstract*— Aggressive behaviour in dementia patients poses significant challenges for caregivers and healthcare providers. This study aims to develop and evaluate Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) models, integrated with word2vec embedding, for accurately predicting aggressive behaviour in dementia patients. Leveraging existing datasets containing pertinent information such as agitation levels and location, our models are trained to discern patterns indicative of aggressive episodes. Healthcare, a complex domain notorious for its diagnostic intricacies, stands to benefit greatly from such predictive analytics. We assess the efficacy of our models by comparing their predictive accuracy against established methodologies in dementia care. Furthermore, we investigate techniques to enhance model performance and discuss potential applications within clinical settings. This research underscores the utility of machine learning and deep learning in addressing critical challenges within healthcare, particularly in the realm of behavioural prediction in dementia care.**

***Keywords*— Healthcare Predictive Analytics, Word2Vec Embedding, Dementia Patients, Machine Learning and Deep Learning.**

1. Introduction

Machine learning (ML) is exerting a growing influence across various sectors, including industry and healthcare. ML applications span from image recognition to personalized product suggestions in online shopping. In healthcare, professionals and researchers are increasingly exploring ML's potential [1]. The integration of artificial intelligence (AI) into medicine dates back to the 1980s and 1990s, with techniques such as fuzzy expert systems, Bayesian networks, artificial neural networks, and hybrid intelligent systems being utilized in diverse medical contexts [2]. The aim has been to enhance medical diagnosis and therapy effectiveness. The evolution of AI in medicine can be categorized into distinct stages. In the 1980s, the development of the "decision tree" algorithm marked a significant milestone, alongside the emergence of artificial neural networks. The 1990s witnessed the continuation of progress with the advent of "expert systems" and support vector machines. By the 2000s, "deep learning" gained prominence, solidifying machine learning as a pivotal aspect of AI in medicine. Presently, the field is in a "maturation period" characterized by advanced technologies, yet challenges persist in interpersonal communication, indicating that we are still in the "weak" AI phase [3].

The limitations of syntactic natural language parsers, particularly their poor performance in domains like the Wall Street Journal, indicate their inadequacy for processing large-vocabulary text with inherent vagueness. Son et al. [4] introduced SPATTER, a statistical parser based on decision-tree learning algorithms, which demonstrates significantly improved accuracy rates compared to existing parsers. SPATTER outperforms grammar-based parsers, such as IBM's computer manuals parser, in various tests. Using PARSEVAL metrics, SPATTER achieves 86 percent precision, 86 percent recall, and 1.3 crossing brackets per sentence for sentences of 40 words or fewer. For sentences spanning 10 to 20 words, SPATTER attains 91 percent precision, 90 percent recall, and 0.5 crossing brackets. Rezvani et al. [5] investigated a methodology aimed at evaluating the likelihood of agitation in dementia patients and the potential assistance offered by in-home monitoring data for such patients. The data utilised in their research was collected during a clinical investigation pertaining to patients with dementia. The proposed model was assessed using a 10-fold cross-validation method, in which the data was divided into separate test and training sets for each fold. The results of the tests, averaged across folds, indicated improved performance in terms of recall, f1-score, and the area under the precision-recall curve for the suggested model, particularly notable in scenarios with unbalanced datasets where accuracy might be misleading. In an unrelated matter, Kumar et al. [6] performed a “systematic literature review” (SLR) to investigate studies that utilised clinical data extracted from electronic health records to simulate the risk of developing “Alzheimer's disease dementia”. They searched various internet databases for publications spanning from 2010 to 2020, selecting relevant articles based on predetermined criteria. AD, the most prevalent cause of dementia characterized by progressive cognitive decline impacting daily functioning, is a neurodegenerative disorder.

Athar et al. [7] proposed a deep learning model utilizing LSTM for predicting agitation episodes in dementia patients. Data were collected from a specialist care facility dedicated to evaluating and treating behavioral and psychological symptoms of dementia, with participants selected from long-term care facilities based on specific inclusion criteria [8]. Consent for the study was obtained from the subjects' authorized decision-makers. Yadav et al. [9] evaluated the efficacy of four supervised machine learning classifiers – Random Forest, Gaussian Naive Bayes, Logistic Regression and Support Vector Machine – for text categorization using datasets such as Reuters, Brown, and movie review corpora. In the study by Khan et al. [10], videos of 20 research participants were

collected for the purpose of conducting the Detecting Agitation study. Researchers examined 35 hours of video footage focusing on individual subjects, demonstrating the utility of video cameras. Shivhare et al. [11] utilized patient interviews with audio recordings, along with transcripts from the largest publicly accessible collection and the Dementia Bank, to train their model for analysis. Zhang et al. [12] examined the utilisation of sensor data and LSTM models to predict human behaviour in smart home environments., exploring optimization techniques such as SGD, Adagrad, Adadelta, RMSProp, and Adam. They employed Word2Vec format for specifying the embedding matrix for each action, achieving an accuracy of 86.32% for activity prediction using the LSTM model.

Muhammad et al. [13] presented a text categorization model that integrates a LSTM network, specifically a variant called "coif-LSTM," with a “convolutional neural network” (CNN) featuring a modified design. Due to the absence of an activation function in the CNN component. Chandola et al. [14] introduced a hybrid “deep learning model” that merges LSTM with Word2vec embeddings. The model underwent testing across various specific businesses spanning diverse industries, including Apple, PepsiCo, NRG, APEI, and AT&T. Luan & Lin [15] proposed a text classification model termed as “NA-CNN-COIF-LSTM”, which integrates CNN and LSTM or a variation thereof with slight modifications. In recent years, there has been a growing interest in leveraging EEG signals for the early diagnosis of Alzheimer's disease, as highlighted by Al-Jumeily et al. [16]. Furthermore, Iram et al. [17] have extended this approach by exploring gait discrimination and neural synchronization for early detection of various neurodegenerative diseases. Additionally, Iram et al. [18] propose a classifier fusion strategy aimed at enhancing the accuracy of early detection methods for neurodegenerative diseases.

2. Materials and Methods

The methodology of this study delineates the approaches employed to acquire, structure, and analyze data, establishing a robust framework for our forthcoming investigations, which will be elaborated upon in subsequent sections.

2.1. Data Collection

The dataset was provided by “Azziza Bankole, MD, Professor of Psychiatry, Program Director, Geriatric Psychiatry Fellowship, Chief Diversity Officer, Virginia Tech Carilion School of Medicine”. To proceed with the task, it's essential to load all necessary dependencies by importing relevant libraries. In Python, the Pandas library can be utilized to obtain a summary of the data. The `describe()` method can offer a quick overview, showcasing the count, standard deviation, mean, maximum, and minimum values for each variable. Additionally, the `info ()` method, as illustrated in Figure 1, can provide a summary encompassing the number of observations, data types, variables, and memory usage. This enables a thorough comprehension of the dataset, including elements such as the quantity of observations, locations, levels, and the span of values for each variable.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 312 entries, 0 to 311
Data columns (total 6 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   timestamp (ET)     312 non-null    object
 1   agitimestamp (ET)  312 non-null    object
 2   Location           312 non-null    object
 3   level              312 non-null    int64
 4   Observation        312 non-null    object
 5   Behaviour          312 non-null    object
dtypes: int64(1), object(5)
memory usage: 14.8+ KB
```

2.2. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial process involving the examination of a dataset to extract insights. It plays a vital role in understanding the data by revealing patterns, trends, and relationships. The primary objective of EDA is to uncover significant insights that can inform subsequent analysis and decision-making processes. To enable numerical operations, particularly in machine learning algorithms, it's necessary to convert the values in the 'Behavior' column from strings to integers, as depicted in Figure 2.

	timestamp (ET)	agitimestamp (ET)	Location	level	Observation	Behaviour
0	12/10/2016 21:02	12/10/2016 21:00	Kitchen	5	Withdrawn	0
1	12/15/2016 3:22	12/15/2016 3:21	Kitchen	3	Vocal1, Withdrawn	0
2	12/15/2016 19:14	12/16/2016 2:20	Other	3	Withdrawn	0
3	12/19/2016 1:57	12/19/2016 1:55	Other	4	Withdrawn	0
4	12/28/2016 20:36	12/28/2016 19:00	Other	4	Vocal2, Withdrawn	0

Figure 2. Converted Values from string to numerical

A column labelled 'word length' was produced (see Figure 3) by applying a Lambda function that calculated the word length in consideration of each space. The results were subsequently represented graphically through the use of a y-axis indicating frequency and an x-axis representing message duration, which was filtered by the terms 'normal' and 'agitation'. The findings of the analysis indicate that the mean word length of reports regarding routine care activities or non-agitative episodes is generally greater in magnitude when compared to reports concerning agitative episodes.

	timestamp (ET)	agitimestamp (ET)	Location	level	Observation	Behaviour	word_length
0	12/10/2016 21:02	12/10/2016 21:00	Kitchen	5	Withdrawn	0	1
1	12/15/2016 3:22	12/15/2016 3:21	Kitchen	3	Vocal1, Withdrawn	0	2
2	12/15/2016 19:14	12/16/2016 2:20	Other	3	Withdrawn	0	1
3	12/19/2016 1:57	12/19/2016 1:55	Other	4	Withdrawn	0	1
4	12/28/2016 20:36	12/28/2016 19:00	Other	4	Vocal2, Withdrawn	0	2

Figure 3. Used Lambda function to calculate the word length

By employing a criterion-specific process to exclude data sets from a data frame, such as mandating the inclusion of a minimum of 10 observations (as illustrated in Figure 4), the precision and dependability of any forecasts or analyses performed on the dataset are guaranteed. This process helps in maintaining the quality of the data and enhances the robustness of subsequent analyses.

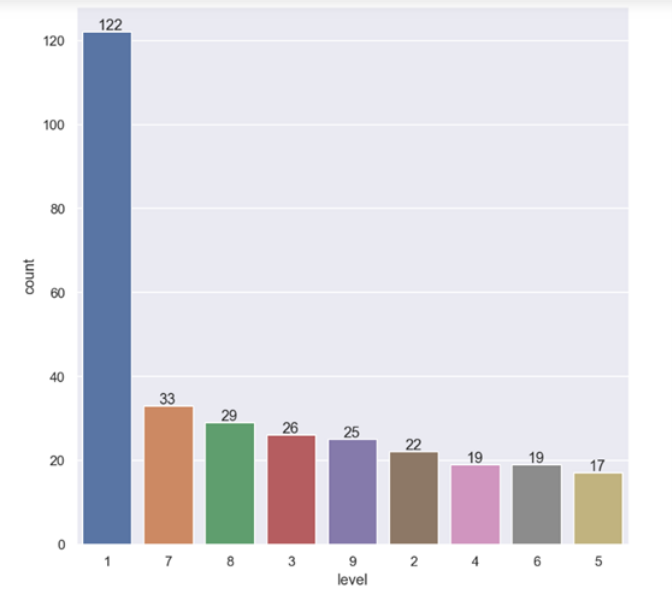

Figure 4. Group by chart

The pandas library was utilized to construct a data frame representing the frequency distribution of words, followed by the utilization of the seaborn library to produce a bar plot (refer to Figure 5). Analyzing the frequency distribution of words within a text dataset serves various purposes: identifying the most prevalent words, discerning topics, and uncovering potential relationships between words. This analysis aids in understanding the context of the text, identifying patterns, and detecting sentiment, common words, and biases within the text. It contributes significantly to text analysis and natural language processing tasks.

3. Methodology

A viable approach involves constructing a model aimed at predicting agitative behavior using text data derived from surveys provided by caregivers. Subsequently, the efficacy of the model can be evaluated through traditional data modeling techniques.

3.1 Data Preprocessing

An essential step in readying data for “natural language processing” (NLP) tasks involves text data pre-processing. This process involves cleaning and standardizing the text data to transform it into a format suitable for further analysis. Tasks within text data pre-processing encompass tokenization, applying regular expressions, converting all words to lowercase, removing stopwords, and performing lemmatization. Furthermore, during the pre-processing stage, the text undergoes a conversion into numerical formats, including vector representations and embeddings.

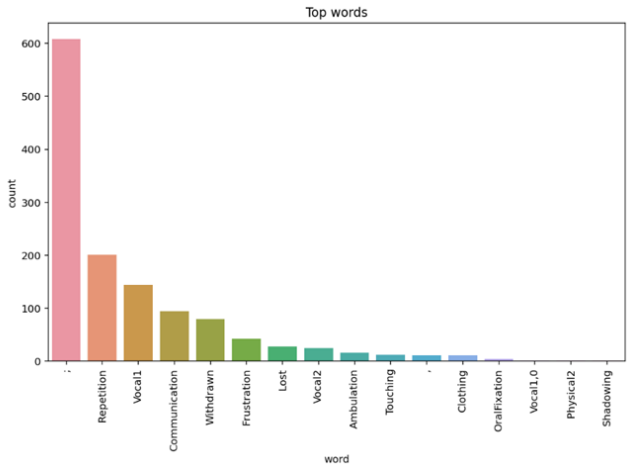

Figure 5. Top words

- Sentence lower: Before the substitution, the operation `sentence. lower()` transforms the sentence to lowercase. For example, if the sentence is "Hello, World!", the regular expression will identify and include the characters ",", and "!", resulting in the modified statement "Hello World".

- Expression (regex): By removing all “non-alphanumeric characters” from a phrase using this regular expression replacement, they are defined as non-numerical characters. The function “re.sub ()” performs a search-and-replace operation on the given sentence, effectively eradicating any matching characters from the sentence by substituting them with an empty string.

- Stop Words Removal: Removing stop words involves creating a new list named "words" using list comprehension in Python. This list comprises all the words from the initial set that are absent in a predetermined list of stopwords. The list comprehension comprises of two components: the iteration that traverses every word in the initial list, and a filter that excludes any word found in the list of stopwords. The result is a new list of words containing only those not identified as stopwords.

- Tokenization: To split a phrase into a list of terms, the `nltk.word_tokenize()` method from the Python “Natural Language Toolkit” (NLTK) package is utilized. This method returns a list of words after receiving a phrase as input, using various approaches

including punctuation and whitespace detection to determine word boundaries. Given an input phrase, such as "Hello, World!", the function will produce a list containing the individual words "Hello" and "World". The generated list, named "words," can be utilised for additional procedures such as word stemming or stopword elimination.

- Lemmatization: Lemmatization is performed by iterating through a list of words called "words" using a loop. The `lemmatize()` method of a lemmatizer object is applied to each word in the list, and the resulting word is added to a new list named "filtered_sentence." This loop use the `lemmatize()` method on each word to generate a new list of words that have been converted to their base form.
-

3.2 Overview of LSTM Model

The LSTM Recurrent Neural Network (RNN) type was initially introduced by Hochreiter and Schmidhuber in 1997. A typical LSTM network structure comprises multiple LSTM cells, each receiving input data (xt) at every iteration (t) and generating an output (ht). In addition to the current cell input (Ct), the LSTM cell also takes into account the previous cell state (Ct-1) and output state (Ct). It incorporates gate mechanisms such as the input gate (it), forget gate (ft), and output gate (ot), which control the flow of information within the network by determining whether certain information should be retained or discarded.

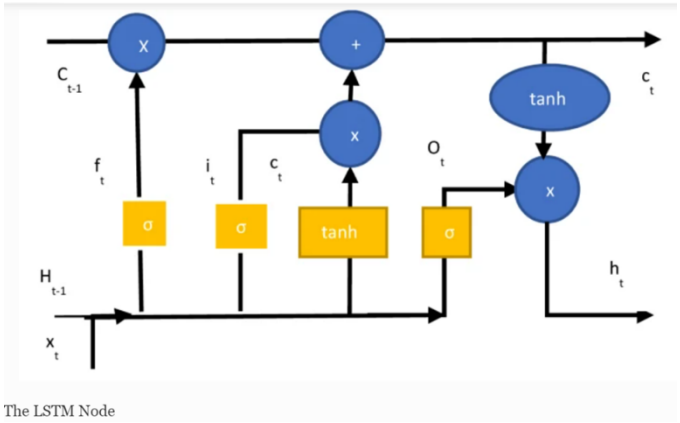

Figure 6. The architecture of LSTM

The most common type of traditional neural networks is the feed-forward neural network. These networks consist of layers of neurons, typically including at least one hidden layer, along with input and output layers. Feed-forward neural networks are limited to static classification tasks, providing a fixed mapping between input and output. For dynamic classification tasks, such as time prediction, a dynamic classifier is required. Although feed-forward neural networks can be enhanced with dynamic classification by feeding signals from previous time steps back into the network, this approach has limitations. RNNs address this issue by incorporating recurrent connections, allowing them to look back in time over multiple steps. However, traditional RNNs face challenges with vanishing or exploding gradients, limiting their ability to capture long-term dependencies. To overcome this, LSTM networks were introduced. LSTMs improve upon traditional RNNs by incorporating memory cells, allowing them to capture long-term dependencies more effectively. Due to their enhanced memory capabilities, LSTM networks can successfully learn from sequences spanning over 1,000 time steps, making them suitable for a wide range of tasks.

3.3 Overview of BILSTM Model

Deep-bidirectional LSTMs enhance the functionality of typical LSTM models by utilising two LSTMs to analyse input data in both forward and backwards directions. At first, an LSTM network analyses the input pattern in the forward layer, and then it analyses the reverse version of the input sequence in the backwards layer during the next cycle. This dual LSTM approach aims to enhance the model's understanding of long-term dependencies, potentially leading to improved accuracy. By utilizing two LSTMs to process input data, deep-bidirectional LSTMs contribute to enhancing the accuracy of LSTM models. The initial LSTM model evaluates the input pattern in the forward direction, whereas the subsequent LSTM model operates on the reversed version of the input sequence in the opposite direction. The model's ability to perform dual processing allows it to more effectively capture long-term dependencies, leading to enhanced accuracy.

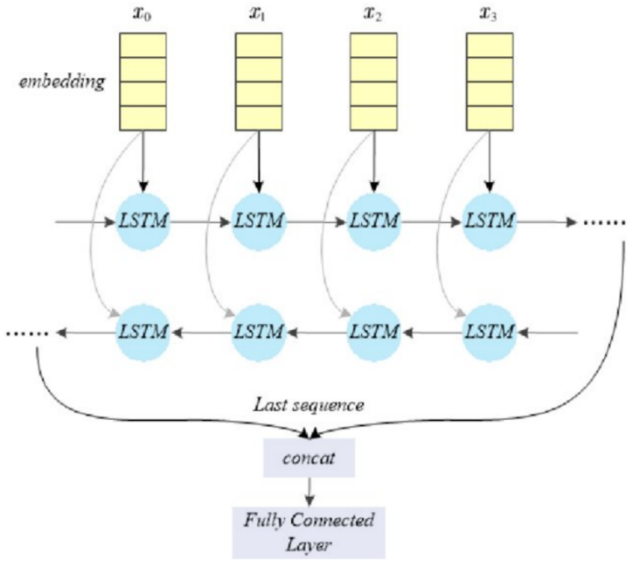

Figure 7. The architecture of BiLSTM

3.4 Overview of WORD2VEC LSTM Model

A novel approach involving a Word2Vec-LSTM model is proposed for predicting aggressive behavior in dementia patients. This model utilizes the Word2Vec layer network to extract semantic information from words, generating text vectors which serve as inputs for the LSTM layer network. The output vectors from the LSTM layer are then passed through a

fully connected layer to extract features from the text. Subsequently, the sigmoid activation function is employed to classify the behavior of dementia patients as either normal or aggressive.

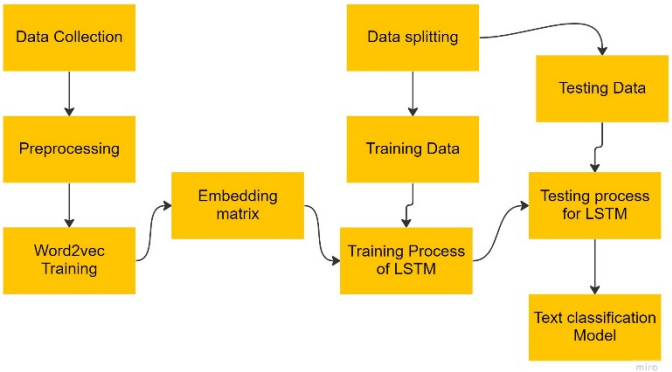

Figure 8. The architecture of WORD2VEC LSTM

3.5 Overview of Word2Vec Model

The Word2Vec model, derived from word embeddings, is a primary method for assessing the similarity and significance of word meanings. It transforms words into vectors, representing them in a continuous vector space. These word vectors produced by the Word2Vec model can be leveraged to compute similarity values using the Cosine Similarity formula. Various parameters, such as window size and vector dimensions, are essential in configuring the Word2Vec model during its training process.

In the training phase of the Word2Vec model, the entire dataset is provided alongside the desired number of training epochs, particularly when employing the Skip-Gram approach. This method involves normalizing text data into pairs of words and their respective contexts (w,c), which are then inputted into a neural network for further processing.

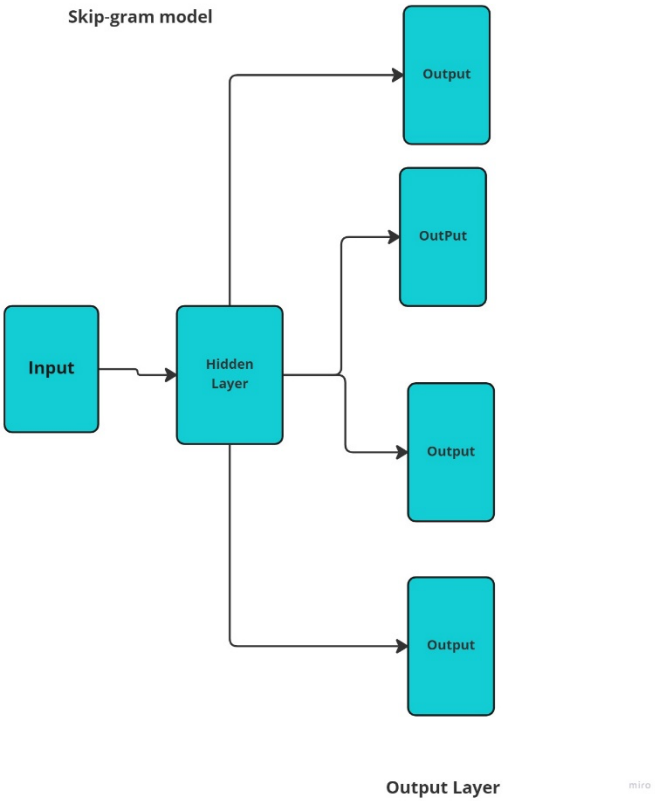

Figure 9. Word2vec Skip-Gram Model

4. Result Analysis

In this section, all important results are given.

4.1 By LSTM Model

The LSTM model, trained over 15 epochs with a batch size of 32, has attained an accuracy of 84.51% on the validation set. This achievement is noteworthy, especially considering the relatively modest size of the training dataset. Throughout the training process, both the training and validation loss consistently decreased with increasing epochs (refer to Figure 10), indicating that the model effectively learns and generalizes from the data. The peak validation accuracy reached during training was 84.913% (refer to Figure 11), suggesting that the model has reached its optimal performance. Overall, these results demonstrate the model's strong performance and its ability to generalize effectively with a high accuracy.

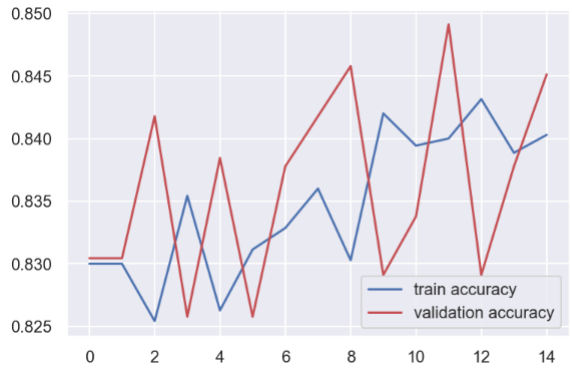

Figure 10. Train validation accuracy

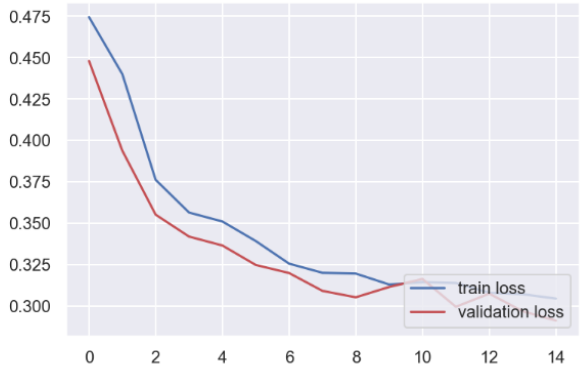

Figure 11. Loss Plot

To enhance the model's performance, several strategies could be considered. Expanding the training dataset or implementing more sophisticated methods such as hyperparameter tuning are possible approaches for enhancing performance. Additionally, optimizing the dropout layer's parameter settings could lead to better accuracy. Notably, in this model, the highest accuracy was observed when the dropout layer was configured with a dropout rate of 0.4 (refer to Figure 12). Adjusting dropout rates and exploring other regularization techniques could further refine the model's performance and contribute to improved accuracy.

	precision	recall	f1-score	support
Agitation	0.56	0.40	0.47	254
Normal	0.88	0.94	0.91	1244
accuracy			0.85	1498
macro avg	0.72	0.67	0.69	1498
weighted avg	0.83	0.85	0.83	1498

Figure 12. LSTM Classification Report

The classification report reveals that the model's performance is suboptimal, particularly for the "Agitation" class. The precision, recall, and f1-score for this class are relatively low, with values of 0.56, 0.40, and 0.47, respectively. In contrast, the model demonstrates strong performance for the "Normal" class, achieving precision, recall, and f1-score values of 0.88, 0.94, and 0.91, respectively. Despite an overall accuracy of 0.85, the macro average of precision and recall indicates that the model's performance is lacking in both classes. This discrepancy could be attributed to class imbalance within the dataset, where the "Normal" class contains significantly more samples than the "Agitation" class (refer to Figure 13). Addressing this class imbalance issue through techniques such as oversampling or undersampling may help improve the model's performance for the "Agitation" class.

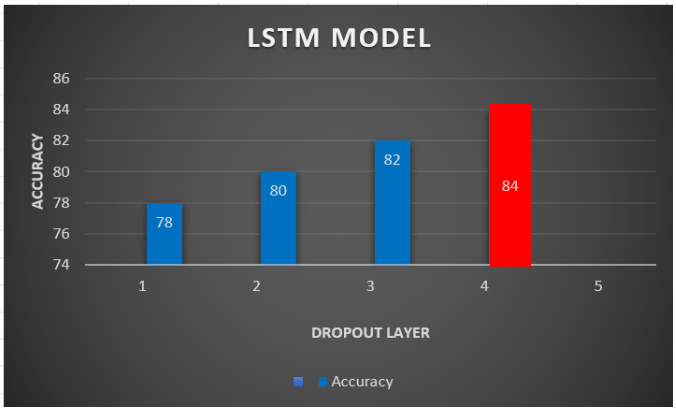

Figure 13. Dropout Layer

4.2 By BiLSTM Model

Training for 15 epochs with a batch size of 35 was performed on the BILSTM model, which was then assessed utilising both the training and validation data. The model demonstrated a validation accuracy of 84.05% and a training accuracy of 84.54%. It is worth noting that the highest validation accuracy of 84.85% was achieved in the fourteenth epoch. A declining trend was observed in both the training and validation loss as the number of epochs increased during the training process. This suggests that the model successfully learned from the training set and was able to generalise well to the validation set. Nevertheless, it is important to acknowledge that the validation accuracy attained a plateau following the fourteenth epoch. This indicates that the model's performance had already been optimised, and additional training iterations might not result in substantial enhancements. In general, the BILSTM model exhibited strong performance, attaining an accuracy of 84.05% when evaluated on the validation set. Notably, adjusting the dropout layer to 0.4 resulted in the highest accuracy for this model (refer to Figure 14).

	precision	recall	f1-score	support
Agitation	0.67	0.13	0.22	256
Normal	0.85	0.99	0.91	1242
accuracy			0.84	1498
macro avg	0.76	0.56	0.57	1498
weighted avg	0.82	0.84	0.79	1498

Figure 14. BiLSTM Classification Report

The macro average was obtained by averaging the metric for each class, while the weighted average took into consideration the support (number of instances) of each class (refer to Figure 15).

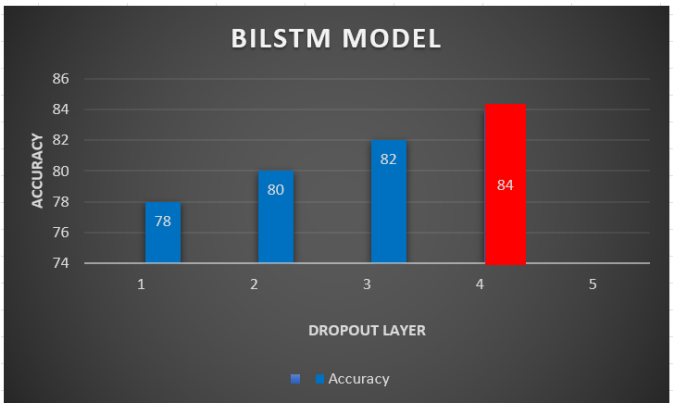

Figure 15. Dropout layer of BiLSTM Classification

4.3 By Word2vec_LSTM Model

The training loss, as determined by the initial epoch, is 0.5191, while the accuracy is 0.8299. The validation loss, meanwhile, is 0.4330, whereas the accuracy is 0.8468. It is noteworthy that the model demonstrates a marginal enhancement when applied to the validation dataset as opposed to the training dataset. Subsequently, the training loss decreases to 0.4584, whereas the accuracy remains unchanged at 0.8299. Concurrently, the validation loss exhibits stability at 0.4285, while the accuracy stands at 0.8468. (refer to Figure 16 and Figure 17). This suggests that the model does not exhibit improvement during the second epoch, as both the training and validation metrics remain unchanged from the first epoch.

The study emphasises that the model attains an overall accuracy of 0.82, indicating a relatively good level of performance. Nevertheless, the global average of precision and recall is quite low, suggesting inadequate performance in both classes. This phenomenon can be ascribed to the presence of an unequal distribution of classes in the dataset or the inadequacy of the data complexity for the given task. Significantly, the model demonstrates an accuracy, recall, and f1-score of 0.00 for the 'Agitation' class, indicating its inability to make precise predictions for this particular class. In contrast, the model achieves a precision of 0.82, recall of 1.00, and f1-score of 0.90 for the 'Normal' class, indicating highly accurate predictions (see Figure 18).

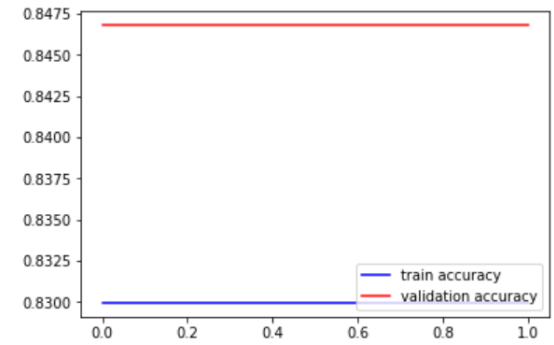

Figure 16. Word2vec_LSTM Accuracy Plot

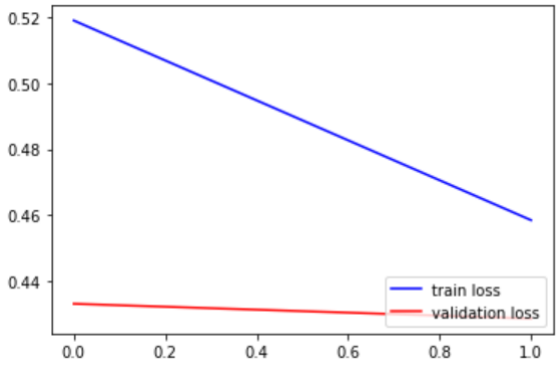

Figure 17. Loss Plot

	precision	recall	f1-score	support
Normal	0.82	1.00	0.90	2714
Agitation	0.00	0.00	0.00	581
accuracy			0.82	3295
macro avg	0.41	0.50	0.45	3295
weighted avg	0.68	0.82	0.74	3295

Figure 18. Word2vec LSTM Classification Report

The model has attained its highest level of accuracy when the dropout layer has been adjusted to 0.4, as seen in Figure 19.

4.1 Word2vec Model

Based on the cosine similarity among their word vectors, this technique's line of code retrieves words that bear the closest resemblance to the input word. The word with the highest similarity score is presented first, indicating its likeness to the input word. In this instance, the code returned words closely related to the term "Withdrawn." These words received high similarity scores from the model, suggesting comparable semantic meanings and usage contexts within the dataset.

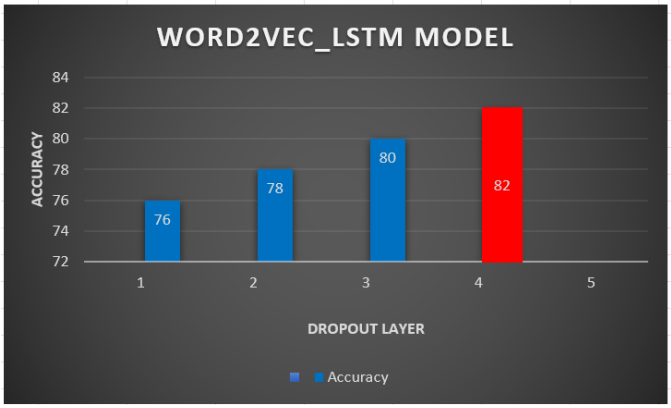

Figure 19. Dropout Layer

4.2 Comparative Result Analysis

- **Normal Class**

According to the table, the LSTM model boasts the highest accuracy rate (85%) and precision (88%). Meanwhile, the word2vec LSTM model achieves the highest F1 score (90%), and the BiLSTM model demonstrates the highest recall rate (99%). Overall, the table illustrates that the LSTM model excels in terms of accuracy and precision, whereas the BiLSTM model outperforms in terms of recall. However, the word2vec LSTM model exhibits a balanced performance in both recall and accuracy (refer to Table 1).

Model	Accuracy	Precision	Recall	F1 Score
LSTM	85%	88%	94%	91%
Bi LSTM	84%	85%	99%	91%
word2vec_LSTM	82%	82%	100%	90%

Table 1. Normal Class

- **Agitation Class**

As per the table, the LSTM model attains the highest accuracy (56%) but exhibits the lowest recall (40%). On the other hand, the BILSTM model records the lowest recall (13%) and the second-highest accuracy (67%). However, the word2vec LSTM model displays the poorest performance with 0% accuracy and recall, indicating its inability to accurately predict the Agitation class. Notably, all models showcase relatively low F1-Scores (refer to Table 2).

Model	Precision	Recall	F1 Score
LSTM	56%	40%	47%
Bi LSTM	67%	13%	22%
word2vec_LSTM	00%	00%	00%

Table 2. Agitation Class

5. Conclusion

This study aims to evaluate the efficacy of four text categorisation techniques in predicting violent behaviour in people with dementia. The initial three models, comprising of a basic LSTM, BILSTM, and word2vec_LSTM, were deep learning models, whereas the fourth model, word2vec, was an unsupervised model. The results demonstrated that all deep learning models exhibited accuracy rates between 82% and 86% on the validation set, indicating commendable performance and generalisation capability. Upon analysing the classification report for the "Agitation" and "Normal" classes, it became evident that the models did not achieve ideal performance for the "Agitation" class. The precision and recall values for this class were relatively poor, suggesting challenges in reliably detecting instances of "Agitation." This can be ascribed to the dataset's class imbalance, as the "Normal" class exhibited a substantially greater number of samples compared to the "Agitation" class. The word2vec model exhibited a word similarity of 92% when assessed using cosine similarity. This method effectively detected words in the dataset that have the same meanings and are used in similar situations, providing valuable information about the similarity of terms and their usage in different contexts.

Nevertheless, although the deep learning models performed well, the study emphasised the need of tackling dataset imbalance, particularly due to the unequal distribution of samples across the "normal" and "agitation" classes. Moreover, the word2vec LSTM hybrid model encountered difficulties in reliably predicting occurrences belonging to the "agitation" class. It is important to consider this constraint when analysing the results of the investigation.

5.1 Limitation of the study

The study was dependent on the accuracy of the data obtained from the intelligent nursing home, which was susceptible to potential errors or inaccuracies. Additionally, the investigation solely focused on the four machine learning algorithms utilized, without considering other potential algorithms or techniques that could have been employed to predict behavior. Furthermore, the small sample size in this study may have influenced the predictive capabilities of the model. An additional constraint was the extremely unbalanced dataset utilised, in which one class possessed a considerably greater quantity of samples compared to the other class. The aforementioned disparity may have had an adverse impact on the efficacy of the predictive models and the overall precision of the results.

5.2 Future work

Future research in this field should focus on improving the effectiveness of text categorisation techniques for forecasting aggressive behaviour in individuals with dementia. One way to accomplish this is by increasing the size of the training dataset

and utilising sophisticated methods like hyperparameter tweaking and different topologies like transformer models. Moreover, acquiring a wider range of data from several sources and rectifying disparities in social class can lead to enhanced results. In order to improve the precision and effectiveness of these models, future research should prioritise the identification of the most suitable set of features for differentiating between periods of agitation and non-agitation. Incorporating carer perspectives into the analysis, creating notifications to inform carers about possible agitation, and providing personalised behaviour management recommendations based on individual patient assessments, including interests, music preferences, activity levels, and mood, are worthwhile areas to investigate.

References

[1] Y. Verma and S. Tayeb, "Evaluation of machine learning architectures in healthcare.," 11th IEEE Annual Computing and Communication Workshop and Conference, CCWC 2021, Las Vegas, NV, USA, January 27-30, 2021.

[2] Amisha, P. Malik, M. Pathaniaand V. K. Rathaur, "Overview of artificial intelligence in medicine", vol. 8, no. 7, Jul. 2019.

[3] R. Liu, Y. Rongand Z. Peng, "A review of medical artificial intelligence", vol. 4, no. 2, May 2020.

[4] J. Son and S. B. Kim, "Rule Selection Method in Decision Tree Models," Journal of Korean Institute of Industrial Engineers, vol. 40, no. 4, pp. 375–381, Aug. 2014.

[5] R. Rezvani, Samaneh Kouchaki, Ramin Nilforooshan, D. J. Sharp, and Payam Barnaghi, "Semi-supervised Learning for Identifying the Likelihood of Agitation in People with Dementia," arXiv (Cornell University), May 2021.

[6] S. Kumar, I. Oh, S. Schindler, A. M. Lai, P. R. O. Payne, and A. Gupta, "Machine learning for modeling the progression of Alzheimer disease dementia using clinical data: a systematic literature review," JAMIA Open, vol. 4, no. 3, Jul. 2021, doi: https://doi.org/10.1093/jamiaopen/ooab052.

[7] S. HekmatiAthar, H. Goins, R. Samuel, G. Byfield, and M. Anwar, "Data-Driven Forecasting of Agitation for Persons with Dementia: A Deep Learning-Based Approach," SN Computer Science, vol. 2, no. 4, Jun. 2021, doi: https://doi.org/10.1007/s42979-021-00708-3.

[8] A. Iaboni et al., "Wearable multimodal sensors for the detection of behavioral and psychological symptoms of dementia using personalized machine learning models," Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, vol. 14, no. 1, Jan. 2022, doi: https://doi.org/10.1002/dad2.12305.

[9] B. P. Yadav, S. Ghate, A. Harshavardhan, G. Jhansi, K. S. Kumar, and E. Sudarshan, "Text categorization Performance examination Using Machine Learning Algorithms," IOP Conference Series: Materials Science and Engineering, vol. 981, p. 022044, Dec. 2020, doi: https://doi.org/10.1088/1757-899x/981/2/022044.

[10] S. S. Khan et al., "Unsupervised Deep Learning to Detect Agitation From Videos in People With Dementia," IEEE Access, vol. 10, pp. 10349–10358, Jan. 2022, doi: https://doi.org/10.1109/access.2022.3143990.

[11] Shivhare, N., Rathod, S. and Khan, M.R., 2021, November. Automatic speech analysis of conversations for dementia detection using lstm and gru. In 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA) (pp. 1-7). IEEE.

[12] J. Zhang, Y. Zhu, X. Zhang, M. Ye, J. Yang, Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas," Journal of Hydrology, vol. 561, pp. 918–929, Jun. 2018, doi: https://doi.org/10.1016/j.jhydrol.2018.04.065.

[13] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, Sentiment Analysis Using Word2vec And Long Short-Term Memory (LSTM) For Indonesian Hotel Reviews," Procedia Computer Science, vol. 179, pp. 728–735, 2021, doi: https://doi.org/10.1016/j.procs.2021.01.061.

[14] D. Chandola, A. Mehta, S. Singh, V. A. Tikkiwal, H.Agrawal Forecasting directional movement of stock prices using deep learning. Annals of Data Science. 2023 Oct;10(5):1361-78.

[15] Y. Luan and S. Lin, Research on Text Classification Based on CNN and LSTM, 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Mar. 2019, doi: https://doi.org/10.1109/icaica.2019.8873454.

[16] Al-Jumeily D, Iram S, Vialatte FB, Fergus P, Hussain A. A novel method of early diagnosis of Alzheimer's disease based on EEG signals. The scientific world journal. 2015 Jan 1;2015.

[17] Iram S, Vialatte FB, Qamar MI. Early diagnosis of neurodegenerative diseases from gait discrimination to neural synchronization. InApplied computing in medicine and health 2016 Jan 1 (pp. 1-26). Morgan Kaufmann.

[18] Iram S, Fergus P, Al–Jumeily D, Hussain A, Randles M. A classifier fusion strategy to improve the early detection of neurodegenerative diseases. International Journal of Artificial Intelligence and Soft Computing. 2015 Jan 1;5(1):23-44.