VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i2.6257

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

DEEP BELIEF NETWORK (DBN) IMPLEMENTATION FOR MULTIMODAL CLASSIFICATION OF SENTIMENT ANALYSIS

Hilmi Hibatullah[1*]; Aris Thobirin[2]; Sugiyarto Surono[3]

Mathematics[1, 2, 3]
Ahmad Dahlan University, Yogyakarta, Indonesia[1, 2, 3]
https://uad.ac.id/[1, 2, 3]
hilmi2100015026@webmail.uad,ac.id[1*], aris.thobi@math.uad.ac.id[2], sugiyarto@math.uad.ac.id[3]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *In sentiment analysis, the use of multimodal data, consisting of a combination of images and text, is becoming increasingly important for understanding digital context. However, the main challenge lies in effectively integrating these two types of data into a single learning model. Deep Belief Network (DBN), with its capability to learn hierarchical data representations, is utilized to explore optimal strategies for multimodal sentiment analysis. The dataset includes 34,034 images from the FERPlus dataset to train the model in classifying emotions based on facial expressions, as well as 999 text and image samples obtained through crawling X. Experiments were conducted by comparing the performance of DBN with 2, 3, and 4 hidden layers across different test data sizes (10%-50%). The results indicate that the 3-hidden-layer configuration achieved the best performance, with a highest accuracy of 76% at a 20% test data size. Additionally, testing different learning rates (10^{-4} to 10^{-7}) produced consistent results, but the fastest computation time was achieved with a learning rate of 10^{-4}. Based on these findings, DBN with a 3-hidden-layer configuration and a learning rate of 10^{-4} is considered a more efficient alternative for multimodal sentiment analysis based on text and images.*

Keywords: *deep belief network, multimodal, sentiment analysis.*

Intisari— *Dalam analisis sentimen, penggunaan data multimodal yang terdiri dari kombinasi gambar dan teks menjadi semakin penting untuk memahami konteks digital. Namun, tantangan utama terletak pada bagaimana menggabungkan kedua jenis data ini dengan baik dalam satu model pembelajaran. Deep Belief Network (DBN), dengan kemampuannya dalam mempelajari representasi hierarkis data digunakan untuk mengeksplorasi strategi optimal dalam menganalisis sentimen berbasis multimodal. Data yang digunakan mencakup 34.034 gambar dari dataset FERPlus untuk melatih model dalam mengklasifikasikan emosi berdasarkan ekspresi wajah, serta 999 teks dan gambar yang diambil melalui crawling X. Percobaan dilakukan dengan membandingkan kinerja DBN pada konfigurasi lapisan tersembunyi 2, 3, dan 4 dengan berbagai ukuran data uji (10%-50%). Hasil penelitian menunjukkan bahwa konfigurasi 3 lapisan tersembunyi memberikan kinerja terbaik, dengan akurasi tertinggi sebesar 76% pada ukuran data uji 20%. Selain itu, pengujian terhadap nilai learning rate (10^{-4} hingga 10^{-7}) menunjukkan hasil yang konsisten, tetapi waktu komputasi tercepat tercapai pada learning rate 10^{-4}. Berdasarkan hasil ini, DBN dengan konfigurasi 3 lapisan tersembunyi dan learning rate 10^{-4} menjadi alternatif yang lebih efisien untuk analisis sentimen multimodal berbasis teks dan gambar.*

Kata Kunci: *deep belief network, multimodal, analisis sentimen.*

INTRODUCTION

Artificial Intelligence (AI) has become one of the fastest-growing fields in this era. Starting from the basics of mathematics such as linear algebra, probability theory, and optimization [1]. AI has developed rapidly and is applied in various fields, including image classification, text classification, speech recognition, computer vision, and others, showing extraordinary potential in increasing

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.6257

efficiency in multiple sectors [2]. One part of AI is sentiment analysis which refers to emotions, views, or attitudes expressed towards a particular thing [3].

Sentiment analysis is the process of detecting the contextual polarity of a text. This analysis determines whether the text is positive, negative, or neutral [4][5]. Due to the rapid development of information technology, social media platforms such as X, Instagram, and Metta have become important components of modern life. Social media platforms not only allow users to exchange thoughts and opinions, but also become a tool for spreading various types of content, including text, images, videos, and audio [6]. Among these platforms, X remains one of the most popular because of its ability to facilitate real-time discussions on various topics [7][8]. In addition to serving as a space for personal interaction, X provides a valuable resource for researchers to study public sentiment and opinion on a large scale [9].

In the field of sentiment analysis, traditional methods generally only process text data, which results in limitations in understanding the overall sentiment context, for example, in a X tweet that includes text and images, the text itself may not be enough to describe the entire emotion or meaning intended by the sender. This causes limitations in sentiment analysis if it does not consider other modalities [10]. Although initially focused only on text, sentiment analysis has now evolved to include multimodal data, which combines text and images to understand the context in more depth about the context and sentiment contained therein [11]. This multimodal approach is increasingly relevant along with the increasing diversity of content uploaded by users on social media, where text is often accompanied by images or even videos that complement each other in conveying messages. However, multimodal integration presents its challenges. Text and images have different data structures, requiring sophisticated models to effectively combine these two types of data into a unified analysis framework [12]. To overcome this challenge, a method that is able to learn complex relationships between modalities is needed, such as Deep Belief Network (DBN). DBN has been shown to be able to handle high-dimensional data and extract meaningful features from complex datasets. Its ability to learn hierarchical representations makes it well-suited for multimodal tasks, as it can understand both textual meaning and visual features in one go. In addition, the layered structure of DBN allows the model to identify relationships within a single modality as well as across modalities, making it an effective method in text- and image-based sentiment analysis [13].

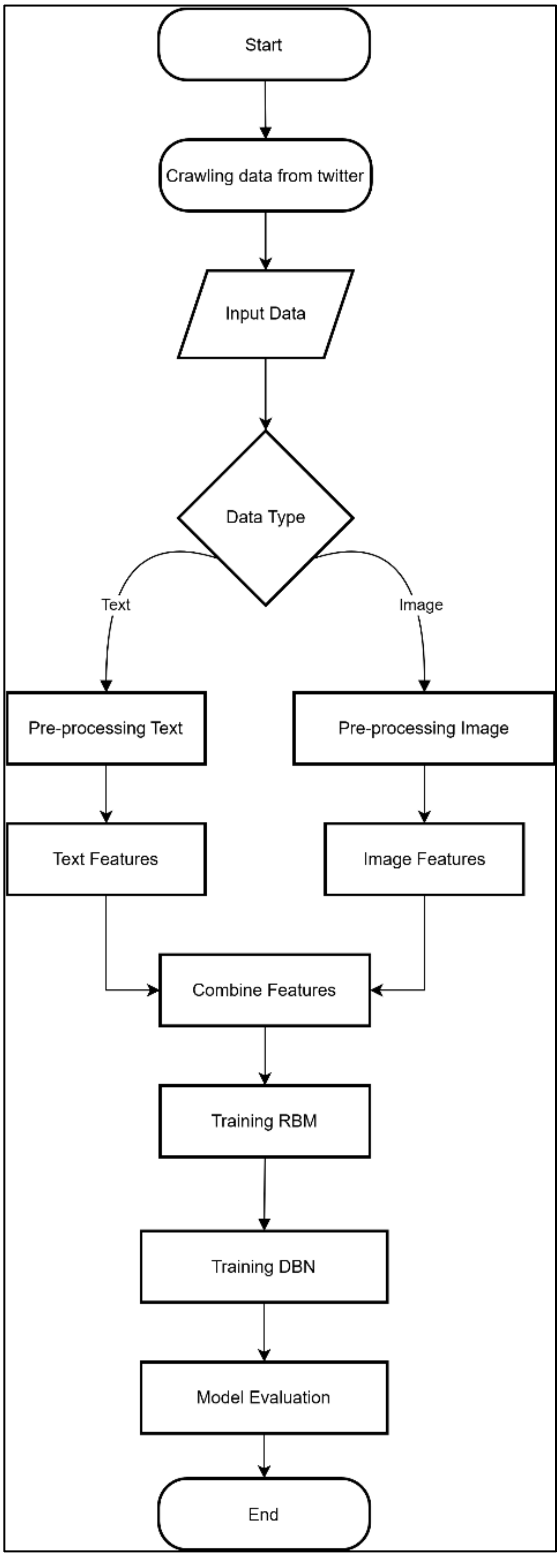
Several previous studies have discussed the use of multimodal approaches in sentiment analysis. Al-Tameemi [10] reviewed text and image based sentiment analysis using feature based methods, while Gandh [11] provided a systematic review of multimodal fusion datasets and techniques but without in depth exploration of deep learning based models. Maurya and Jha [12] proposed a hybrid approach but did not explore the optimization of data integration in depth. This study addresses these gaps by applying DBN to effectively combine text and image data, evaluating various hidden layer configurations, and optimizing the learning rate to determine the best combination for sentiment classification.

DBN demonstrates advantages in capturing complex patterns across various contexts. In the study by Song [14], DBN proved effective in financial data analysis, enhancing market prediction accuracy by extracting features from large and diverse datasets using hierarchical layers of Restricted Boltzmann Machines (RBM). Scarpiniti [15] highlight DBN's capability in classifying sounds at construction sites, where the model can identify sound patterns from various construction tools, showcasing its effectiveness in handling multimodal data. Additionally, Sohn [16] emphasizes that DBN is a powerful technique for intrusion detection, capable of capturing patterns in complex and diverse network data and demonstrating superiority over traditional machine learning methods. Lastly, Zeng [17] employed DBN in Alzheimer's diagnosis, utilizing multimodal data to identify relevant features from MRI and psychological data, thereby improving classification accuracy across different stages of the disease. Thus, the advantages of DBN in capturing complex patterns and handling multimodal data are highly relevant to your research in multimodal sentiment analysis classification involving both text and images.

Unlike previous approaches that focus more on conventional multimodal fusion techniques, this study explores the effectiveness of DBN in learning complex hierarchical representations from text and image data. The layered structure of DBN allows for automatic feature extraction, resulting in a more comprehensive sentiment analysis. By evaluating various configurations of hidden layers and learning rates, this study provides important insights into optimizing DBN for multimodal sentiment classification.

MATERIALS AND METHODS

The stages of research in analyzing sentiment from X multimodal data are outlined in Figure 1.



Source : (Research Results, 2024)

Figure 1. Research Stage

Data Collection

This study utilizes data from X for multimodal sentiment analysis, incorporating both text and images. The image data features pre-labeled facial expressions sourced from the FERPlus dataset, while the text data consists of tweets collected through web crawling. By combining visual and textual features from both sources, the study aims to capture the dynamics of sentiment more effectively.

The dataset used in this study consists of two main sources, namely image data and text data, which allow for a multimodal sentiment analysis approach. The image data includes 34.034 labelled facial expressions from FERPlus data, which are used to label 999 image data resulting from X crawling. Meanwhile, the text data is in the form of tweets related to user emotions collected from January 2023 to October 2024, using keywords and hashtags such as "happy", "sad", "angry", "scared", and others. This dataset is designed to categorize sentiment into three main classes: positive (69.5%), negative (26.4%), and neutral (4.1%), which are the main foundations of the sentiment analysis and classification process in this study.

To evaluate the impact of different training and testing proportions on model performance, five dataset splits were applied: 90% training - 10% testing, 80% training - 20% testing, 70% training - 30% testing, 60% training - 40% testing, and 50% training - 50% testing. These variations were tested to determine the optimal data distribution for achieving the best balance between model training and generalization in multimodal sentiment classification using DBN. Table 1 summarizes the sources and descriptions of the datasets used in this study.

Table 1. Data Source

No	Data Source	Description
1	FERPlus Dataset	34.034 labeled facial expressions for various emotions
2	X Crawling Result Dataset	999 tweets related to user emotions from January 2023 to October 2024

Source : (Research Results, 2024)

Preprocessing Data

The data preprocessing stage aims to prepare the data so that it is ready to be used in further analysis.

1. Preprocessing Teks

In text data, the preprocessing steps include :

a. Text cleaning : is an important stage in Natural Language Processing (NLP) that aims to improve the quality and reliability of text data by

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.6257

converting text to lowercase for consistency and removing URLs, mentions, hastags, and non-alphabetic characters [18].

b. Tokenizing : the process of separating sentences into their constituent words. Here, punctuation characters are divided and separated into independent tokens [19].
c. Stopword Removal : removing common words that do not have a significant contribution to meaning [20], such as “and”, “at”, “which”, or “is”.
d. Stemming : Convert words into their base form by removing prefixes, insertions, suffixes, and combinations of prefixes and suffixes to reduce redundancy [21].
e. TF-IDF Weighting : a text representation method used to evaluate the importance of a word to a document. Term Frequency (TF) represents word frequency, namely how often the words appear in the corpus. Meanwhile, Inverse Document Frequency (IDF) measures the importance of a term in the entire corpus. The TF-IDF weight is calculated by multiplying the two measures. Thus, the higher the weight, the more significant the word in question is in the corpus [22].

2. Image Preprocessing

In image data, preprocessing is done by pixel normalization and resizing the image to standard dimensions to match the model input and improve model performance [23]. In addition, image features are extracted using the Convolutional Neural Network (CNN) model which provides excellent results in the fields of pattern recognition, and image processing, and can extract features from data through convolution [24][25]. Text and image features are combined into a unified feature vector to represent the overall multimodal data:

$$Combine_{features} = [Text_{features}, Image_{features}] \quad (1)$$

After the two features are combined, the combined features are then entered into the DBN model.

Experiment and Modelling

At this stage, experiments are conducted to train and evaluate the Deep Belief Network (DBN) model using the combined text and image features. DBN consists of several Restricted Boltzmann Machines (RBMs) that are trained sequentially. The training process starts from the lower layer RBM using training data, where the resulting representation is then used as input data to train the RBM in the next layer. This process continues until all layers have been trained [26]. The combined feature vector of text and images is entered into a DBN with 2, 3, and 4 hidden layers. Each hidden layer is trained as an RBM, with the following energy function:

$$E(v,h) = -b^T v - c^T h - v^T W h \quad (2)$$

Where v is the visible unit, h is the hidden unit, W is the weight between v and h, and b and c are the biases for each unit.

After training, the probability of a hidden unit being 1 can be expressed as:

$$P\left(h_j = 1 \middle| v\right) = \sigma(b_j + \textstyle\sum_i v_i W_{ij}) \quad (3)$$

The probability of a unit being seen with a value of 1 can be expressed as:

$$P(v_i = 1|h) = \sigma(b_i + \textstyle\sum_j h_j W_{ij}) \quad (4)$$

The DBN model utilizes the output of the previous RBM layer as input for the next layer until all layers are trained.

Model Evaluation

Model evaluation is performed using precision, recall, f1-score, and accuracy metrics to assess the performance of multimodal sentiment classification applied to image and text data. Accuracy is used to measure the proportion of correct predictions from the entire data, while precision evaluates the extent to which predictions are truly relevant to a particular class. Recall measures the extent to which the model can recognize all positive data in the class, and the F1-score provides a balance between precision and recall, especially when there is data imbalance between classes [19].

RESULTS AND DISCUSSION

The data used in this study consists of two sources, image data from the FERPlus dataset and text and image data taken through crawling on X. The image dataset from FERPlus is used to train the model in classifying emotions based on facial expressions, while image data taken from X is combined with text data for multimodal sentiment analysis.

This experiment was conducted by implementing the Deep Belief Network (DBN) using three configurations of the number of hidden layers, namely 2, 3, and 4. The main objective of this study is to determine the best configuration that provides optimal performance in multimodal sentiment classification where text and image features are combined.

The 2 hidden layer configuration was used as a baseline to understand the model's performance with a simpler architecture. To improve performance, the number of hidden layers was increased to 3, and the results showed a more balanced improvement compared to 2 hidden layers. Next, testing was conducted with 4 hidden layers to explore potential further improvements. However, the results indicated that the accuracy obtained was similar to that of 3 hidden layers, while execution time increased. This suggests that adding more layers did not provide significant improvements in evaluation performance. Since 4 hidden layers already produced the same results as 3 hidden layers, configurations beyond this were not used.

The different test sizes (10% to 50%) evaluate model performance at different levels. 10% was used for initial testing with limited data, while 20%-30% provided a balanced ratio for stable and representative results. 50% enabled a comprehensive evaluation of the model's ability to recognize patterns and generalize. These variations allowed an in-depth analysis of DBN performance and the trade-off between model complexity and data availability.

Testing with 2 Hidden Layer

In the first test using 2 hidden layers, the model was tested with various sizes of test data ranging from 10% to 50% of the total dataset. The detailed evaluation results can be seen in Table 2.

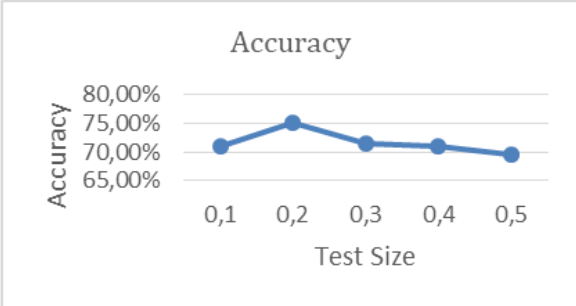
Table 2. Result of 2 Hidden Layers of DBN

Test Size	Time	Precision	Recall	F1-Score	Accuracy
0,1	4'	0,71	1	0,83	0,710
0,2	5'	0,75	0,99	0,86	0,750
0,3	5'	0,72	0,99	0,83	0,713
0,4	6'	0,71	1	0,83	0,710
0,5	6'	0,70	0,99	0,82	0,696

Source : (Research Results, 2024)

The results of DBN analysis using these 2 hidden layers show that the model works better on a test data size of 0.2 where the model achieves a precision of 0.75, recall of 0.99, and F1-score 0.86, with the highest accuracy of 0.750. Configuration with 2 hidden layers was used as a baseline to understand model performance with a simpler architecture. This allows an initial evaluation before increasing model complexity to improve performance.

To better illustrate the accuracy trend across different test sizes, Figure 2 presents a visual representation of the accuracy results for 2 hidden layers.



Source : (Research Results, 2024)

Figure 2. Accuracy of 2 Hidden Layers Based on Test Size

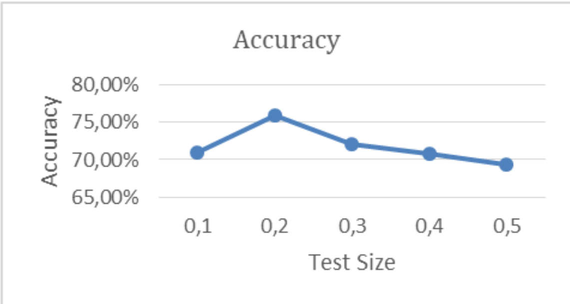
Testing with 3 Hidden Layers

To improve performance, the number of hidden layers was increased to 3, the DBN model showed a significant increase in accuracy. The results obtained are as follows:

Table 3. Results of 3 Hidden Layers DBN

Test Size	Time	Precision	Recall	F1-Score	Accuracy
0,1	5'	0,71	1	0,83	0,710
0,2	5'	0,76	1	0,86	0,760
0,3	6'	0,72	1	0,84	0,720
0,4	7'	0,71	1	0,83	0,707
0,5	7'	0,69	1	0,82	0,694

Source : (Research Results, 2024)



Source : (Research Results, 2024)

Figure 3. Accuracy of 3 Hidden Layers Based on Test Size

In Table 3 and Figure 3, it can be seen that the use of 3 hidden layers produces better performance compared to 2 hidden layers, especially at a test data size of 0.2, where the model achieves the highest precision of 0.76, recall of 1.00, and F1-score of 0.86, with the best accuracy of 0.760.

Testing with 4 Hidden Layers

To explore further performance improvements, testing was conducted with 4 hidden layers showed the same results as 3 hidden layers but different in computing time, the following results were obtained:

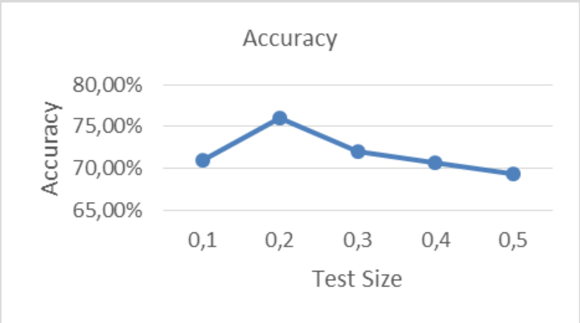
Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.6257

Table 4. Results of 4 Hidden Layers DBN

Test Size	Time	Precision	Recall	F1-Score	Accuracy
0,1	15’	0,71	1	0,83	0,710
0,2	16’	0,76	1	0,86	0,760
0,3	16’	0,72	1	0,84	0,720
0,4	17’	0,71	1	0,83	0,707
0,5	17’	0,69	1	0,82	0,694

Source : (Research Results, 2024)



Source : (Research Results, 2024)
Figure 4. Accuracy of 4 Hidden Layers Based on Test Size

The results in Table 4 and Figure 4 show that the model with 4 hidden layers provides consistent results with the previous configuration. At a test data size of 0.2, the model achieves the best performance with a precision value of 0.76, a recall of 1.00, and an F1-score of 0.86, with an accuracy of 0.760, similar to the results in the 3 hidden layer configuration, while the execution time increased significanly. This indicates that adding more layers does not provide significant improvements in evaluation performance. Since 4 hidden layers already yield the same results as 3 hidden layers, configurations beyond this were not used. Excessive layering often does not significantly improve accuracy but instead increases computation time and the risk of overfitting.

Comparison of Results of 2, 3, and 4 Hidden Layers

Based on the test results, the 3 hidden layer configurations in the DBN model provide optimal results compared to the 2 and 4 hidden layer configurations. This model achieves the best balance between performance and efficiency. In the 3 hidden layer configuration, the highest accuracy is achieved at a test data size of 0.2, with an accuracy value of 0.760. as shown in Figure 5.

In addition, the 3 hidden layer configuration requires a shorter training time compared to 4 hidden layers, without significant performance loss. In contrast, the 2 hidden layer configuration shows good time efficiency but does not provide equivalent results in terms of accuracy and F1-score. This indicates that adding more layers beyond 3 hidden layers does not provide a substantial performance boost.

Source : (Research Results, 2024)
Figure 5. Comparison of Accuracy Results of 2,3, and 4 Hidden Layers

Source : (Research Results, 2024)
Figure 6. Comparison of Computational Time of 2,3, and 4 Hidden Layers

Therefore, as illustrated in Figure 6 the 3 hidden layer configuration is chosen as the optimal configuration because it can provide an ideal combination of high accuracy and computational efficiency.

The Effect of Learning Rate on DBN Performance

After determining that the 3 hidden layer configuration is the most optimal, the next step is to test the effect of the learning rate on model performance. Testing is carried out with various learning rate values, ranging from 10^{-4} to 10^{-7}. Table 5 presents the results obtained:

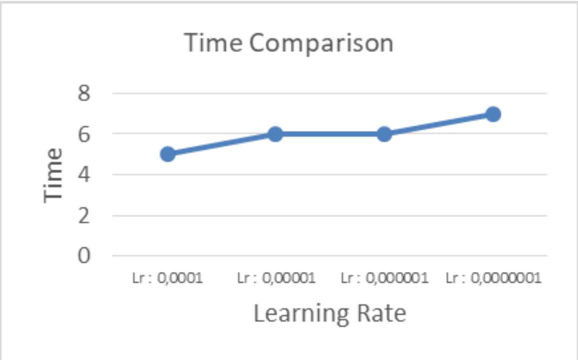
Table 5. Effect of Learning Rate on DBN Performance

Learning rate	Time	Precision	Recall	F1-Score	Accuracy
10^{-4}	5’	0,76	1	0,86	0,76
10^{-5}	6’	0,76	1	0,86	0,76
10^{-6}	6’	0,76	1	0,86	0,76
10^{-7}	7’	0,76	1	0,86	0,76

Source : (Research Results, 2024)

These results show that the precision, recall, F1-score, and accuracy values remain consistent at all tested learning rate values. A learning rate of 10^{-4} produces the fastest and most efficient computation time, without sacrificing the quality of the results. This makes it the best choice to use in this model configuration. The higher computation time speed allows the model to be applied more quickly to large-scale data while maintaining optimal performance.

Furthermore, Figure 7 illustrates a comparison of computation times between learning rates of 10^{-4} and 10^{-7}, emphasizing the efficiency of a higher learning rate.



Source : (Research Results, 2024)

Figure 7. Comparison of Learning Rate Computation Time between 10^{-4} and 10^{-7}

In the conducted by A. Gandhi et al. [11], the SVM method was applied for multimodal sentiment analysis and achieved an accuracy of 70.85%, while the BLSTM approach only reached 65.2%. Compared to the results of this study, the DBN approach with the optimal configuration (3 hidden layers) achieved an accuracy of 76%, which is higher than both methods. This suggests that DBN can effectively capture complex relationships between text and images in multimodal sentiment analysis tasks compared to SVM or BLSTM-based approaches.

However, this study has limitations. The dataset may not fully capture real-world diversity, and model performance could vary with different data. Additionally, DBN's high computational cost may limit scalability. Future research should explore larger datasets and compare DBN with more advanced deep learning models for further validation.

CONCLUSION

In this study, the Deep Belief Network (DBN) was successfully applied to multimodal sentiment analysis using text and image data from X. The model was tested with 2, 3, and 4 hidden layers and various learning rate values. The FERPlus dataset trained the model for emotion classification, while text and image data from X were used for sentiment analysis.

The experimental results show that DBN with 3 hidden layers provides the best balance between accuracy (76%) at a test size of 0.2 and computational efficiency, outperforming 2 hidden layers and matching 4 hidden layers with lower execution time. The optimal learning rate (10^{-4}) ensures faster computation without compromising performance.

However, this study has limitations, including dataset diversity and limited multimodal aspects, as it only combines text and image data. Future research could incorporate larger datasets, as well as explore the integration of additional modalities, such as audio and video and additionally, applying other optimization techniques, such as ensemble methods like Bagging, could enhance the robustness and accuracy of the model in analyzing complex multimodal data.

Overall, this research demonstrates DBN’s effectiveness in multimodal sentiment analysis, it provides a strong foundation for further studies on public sentiment analysis across social media platforms.

REFERENCE

[1] K. Singh, “The Role of Mathematics in Artificial Intelligence and Machine Learning,” *International Journal for Research Publication and Seminar*, vol. 14, no. 5, pp. 186–197, Dec. 2023, doi: 10.36676/jrps.v14.i5.1434.

[2] S. Zhao, F. Blaabjerg, and H. Wang, “An Overview of Artificial Intelligence Applications for Power Electronics,” *IEEE Transactions on Power Electronics*, vol. 36, no. 4, pp. 4633–4658, Apr. 2021, doi: 10.1109/tpel.2020.3024914.

[3] N. Parveen, P. Chakrabarti, B. T. Hung, and A. Shaik, “Twitter sentiment analysis using hybrid gated attention recurrent network,” *Journal of Big Data*, vol. 10, no. 1, Apr. 2023, doi: 10.1186/s40537-023-00726-3.

VOL. 10. NO. 3 FEBRUARY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i3.6257

[4] A. Daza, N. D. González Rueda, M. S. Aguilar Sánchez, W. F. Robles Espíritu, and M. E. Chauca Quiñones, “Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works,” *International Journal of Information Management Data Insights*, vol. 4, no. 2, p. 100267, Nov. 2024, doi: 10.1016/j.jjimei.2024.100267.

[5] P. Subarkah, B. A. Kusuma, and P. Arsi, “SENTIMENT ANALYSIS ON RENEWABLE ENERGY ELECTRIC USING SUPPORT VECTOR MACHINE (SVM) BASED OPTIMIZATION,” *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 2, pp. 252–260, Nov. 2024, doi: 10.33480/jitk.v10i2.5575.

[6] Q. A. Xu, V. Chang, and C. Jayne, “A systematic review of social media-based sentiment analysis: Emerging trends and challenges,” *Decision Analytics Journal*, vol. 3, p. 100073, Jun. 2022, doi: 10.1016/j.dajour.2022.100073.

[7] A. P. Rodrigues et al., “Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–14, Apr. 2022, doi: 10.1155/2022/5211949.

[8] R. A. Rudiyanto and E. B. Setiawan, “Sentiment Analysis Using Convolutional Neural Network (CNN) and Particle Swarm Optimization on Twitter,” *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 2, pp. 188–195, 2024, doi: 10.33480/jitk.v9i2.5201.

[9] Y. Wang, J. Guo, C. Yuan, and B. Li, “Sentiment Analysis of Twitter Data,” *Applied Sciences*, vol. 12, no. 22, p. 11775, Nov. 2022, doi: 10.3390/app122211775.

[10] I. K. S. Al-Tameemi, M.-R. Feizi-Derakhshi, S. Pashazadeh, and M. Asadpour, “A comprehensive review of visual–textual sentiment analysis from social media networks,” *Journal of Computational Social Science*, vol. 7, no. 3, pp. 2767–2838, Sep. 2024, doi: 10.1007/s42001-024-00326-y.

[11] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, and A. Hussain, “Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions,” *Information Fusion*, vol. 91, pp. 424–444, Mar. 2023, doi: 10.1016/j.inffus.2022.09.025.

[12] C. G. Maurya and S. K. Jha, “Sentiment Analysis: A Hybrid Approach on Twitter Data,” *Procedia Computer Science*, vol. 235, pp. 990–999, 2024, doi: 10.1016/j.procs.2024.04.094.

[13] Z. Luo, Z. Tao, Y. Huang, K. He, and C. Liu, “MC-DBN: A Deep Belief Network-Based Model for Modality Completion,” 2024, [Online]. Available: http://arxiv.org/abs/2402.09782

[14] R. Song, Z. Wang, L. Guo, F. Zhao, and Z. Xu, “Deep Belief Networks (DBN) for Financial Time Series Analysis and Market Trends Prediction,” *World Journal of Innovation and Modern Technology*, vol. 7, no. 4, pp. 1–10, Jul. 2024, doi: 10.53469/wjimt.2024.07(04).01.

[15] M. Scarpiniti, F. Colasante, S. Di Tanna, M. Ciancia, Y.-C. Lee, and A. Uncini, “Deep Belief Network based audio classification for construction sites monitoring,” *Expert Systems with Applications*, vol. 177, p. 114839, Sep. 2021, doi: 10.1016/j.eswa.2021.114839.

[16] I. Sohn, “Deep belief network based intrusion detection techniques: A survey,” *Expert Systems with Applications*, vol. 167, p. 114170, Apr. 2021, doi: 10.1016/j.eswa.2020.114170.

[17] N. Zeng, H. Li, and Y. Peng, “A new deep belief network-based multi-task learning for diagnosis of Alzheimer’s disease,” *Neural Computing and Applications*, vol. 35, no. 16, pp. 11599–11610, Jun. 2021, doi: 10.1007/s00521-021-06149-6.

[18] R. B. Hadiprakoso, H. Setiawan, R. N. Yasa, and Girinoto, “Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding,” *AIP Conference Proceedings*, vol. 2680, no. 1, 2023, doi: 10.1063/5.0126116.

[19] T. Ghosh and S. K. B. Math, *: A Concise yet In-Depth Guide on Fundamentals of Computer Vision, NLP, Complex Deep Neural Networks and Machine Learning (English Edition)*. India: BPB publications. 2022.

[20] S. Sarica and J. Luo, “Stopwords in technical language processing,” *PLOS ONE*, vol. 16, no. 8, p. e0254937, Aug. 2021, doi: 10.1371/journal.pone.0254937.

[21] R. Pramana, Debora, J. J. Subroto, A. A. S. Gunawan, and Anderies, “Systematic Literature Review of Stemming and Lemmatization Performance for Sentence Similarity,” *2022 IEEE 7th International Conference on Information Technology and*

Digital Applications (ICITDA), pp. 1–6, Nov. 2022, doi: 10.1109/icitda55840.2022.9971451.

[22] M. Chiny, M. Chihab, O. Bencharef, and Y. Chihab, "LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021, doi: 10.14569/ijacsa.2021.0120730.

[23] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.

[24] S. Surono, M. Y. F. Afitian, A. Setyawan, D. K. Eni Arofah, and A. Thobirin, "Comparison of CNN Classification Model using Machine Learning with Bayesian Optimizer," *HighTech and Innovation Journal*, vol. 4, no. 3, pp. 531–542, Sep. 2023, doi: 10.28991/hij-2023-04-03-05.

[25] S. Surono, M. Rivaldi, D. A. Dewi, and N. Irsalinda, "New Approach to Image Segmentation: U-Net Convolutional Network for Multiresolution CT Image Lung Segmentation," *Emerging Science Journal*, vol. 7, no. 2, pp. 498–506, Feb. 2023, doi: 10.28991/esj-2023-07-02-014.

[26] X. Zhang, N. Gu, J. Chang, and H. Ye, "Predicting stock price movement using a DBN-RNN," *Applied Artificial Intelligence*, vol. 35, no. 12, pp. 876–892, Aug. 2021, doi: 10.1080/08839514.2021.1942520.