Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

DNN-based Drum Performance Analysis System for the Visual Impaired

Jonguk Jung[1], Yunseo Choi[2] and Yoosoo Oh[3]

[1)] Department of Computer and Information Engineering, Daegu University, Gyeongsan 38453, Republic of Korea

[2)] School of AI, Daegu University, Gyeongsan 38453, Republic of Korea

*[3)] School of Computer & Information Engineering, Daegu University, Gyeongsan 38453, Republic of Korea *yoosoo.oh@daegu.ac.kr*

Abstract. In this paper, we propose a drum-style classification system for the visually impaired. The system collects data through the dedicated drumstick and classifies the drum styles. To solve the data imbalance problem, drum style was extracted, and 2000 data were used for rock, hip-hop, funk, and punk. We trained Ensemble voting, RandomForest, DNN, CNN, LSTM, GRU, Ensemble voting with DNN, Ensemble Bagging with DNN, and One-vs-All(OvA) with DNN to compare evaluation performance. As a result, One-vs-All(OvA) with DNN showed the best performance with test accuracy(99.95%). The proposed system contributes to the visually impaired understanding of their musical tastes and improves their self-efficacy.

Keywords: Drum Style, Drum Stick, Visual Impaired, DNN

1. Introduction

According to the 2017 World Health Organization (WHO) statistics, more than 253 million people worldwide have visual impairment[1]. Approximately 0.5% of the global population experiences severe visual impairments and blindness. Despite significant advances in ophthalmology, the number of people affected by these conditions is expected to increase substantially in the coming decades[2]. Visual impairment significantly impacts a person's ability to perform daily activities and engage with their environment, leading to low self-esteem, isolation, depression, and psychological distress[3,4]. Music therapy for visually impaired individuals involves various methods, such as tactile and auditory therapy, instrument playing, and choir activities[5,6,7,8,9]. Music therapy acts as a tool for non-verbal communication, enhancing self-esteem and a sense of accomplishment and contributing to the overall emotional and social development of visually impaired individuals.

Additionally, music therapy effectively improves self-efficacy in visually impaired individuals, allowing them to recognize their abilities through musical achievements and build confidence in facing challenges through gradual progress[10]. Specifically, playing musical instruments helps reinforce self-efficacy by achieving concrete goals[5]. Music therapy using percussion instruments provides opportunities for emotional expression and stress relief, positively influencing self-esteem and emotional stability[7,11]. However, there is currently limited research on using drums among percussion instruments. Therefore, this paper proposes a DNN-based Drum Performance Analysis System for analyzing the drumming style of visually impaired individuals. The proposed system allows visually impaired individuals to arrange virtual drums in desired positions through a virtual drum system. It also classifies the genre of visually impaired individuals' drumming performances using a dataset consisting of ten features, including BPM, performance time (seconds), and the number of times the Crash Cymbal, Small Tom, Middle Tom, Ride Cymbal, High Hat, Snare, Kick, and Floor Tom are played. By verifying whether the classified genre matches the style intended by

* Corresponding author: yoosoo.oh@daegu.ac.kr Received:; Accepted:; Published:

the visually impaired individual, the proposed system aims to help visually impaired individuals feel a sense of accomplishment through their performance, strengthen positive self-perception, and improve self-efficacy.

2. Related Research

2.1 Development of Haptic Wearable Devices for Visually Impaired Individuals

Turchet et al. designed a Music Haptic Wearable (MHW) device for visually impaired individuals, proposing a technology that allows musicians to synchronize through tactile signals without visual cues[12]. In the study by Dishant Zaveri et al., an augmented reality-based virtual drum system called Aero Drums was proposed, enabling drumming in a virtual environment without physical drums using blue sticks[13]. The research by Leon Lu et al. proposed a Music Haptic Wearable (MHWs) system that provides vibration-based tactile feedback for visually impaired music learners, supporting music learning and performance without visual information[14]. Bei Yuan and Eelke Folmer's study introduced a game called Blind Hero for visually impaired individuals. In this study, a haptic glove providing tactile feedback was developed, allowing visually impaired users to play the game[15]. Siddharth Kalra et al. proposed an infrared-based vibration tactile piano learning system for visually impaired individuals to transmit correct piano key and finger information through infrared devices and tactile gloves, enabling piano learning without sight[16]. Based on these studies, this paper proposes a system that supports wireless connectivity, allowing visually impaired individuals to play drumsticks freely, eliminating the inconvenience of complex wired connections, and providing an intuitive drumming experience.

2.2 The Impact of Music Therapy on the Musical Development of Visually Impaired Individuals

Magwati and Manatsa studied how visually impaired individuals express and overcome emotions through various instrumental activities, recognizing music's role as a vital means of expressing emotional responses and processing negative emotions healthily. The sense of achievement experienced by visually impaired individuals through musical accomplishments enhanced personal confidence and social recognition[5]. The research by Petra Kern and Mark Wolery involved designing a program that allowed visually impaired toddlers to independently use six musical stations. Auditory feedback was provided through music to create connections, enabling children to navigate the stations themselves. The results showed that using musical stations increased peer interactions and positively affected play participation while reducing fixed behaviors[6]. Kristi Faby proposed a music therapy program targeting children with Cerebral Visual Impairment (CVI). This study aimed to enhance children's visual attention, sensory efficiency, and social interaction skills through various therapeutic interventions utilizing instruments like drums. It demonstrated that these music therapy interventions positively impacted children's physical, cognitive, and sensory development with CVI [7]. Markus Daube conducted music therapy sessions focused on self-expression through singing and playing instruments. The sessions showed that music therapy could help visually impaired individuals change their negative self-perceptions, enhance self-esteem, and improve social interaction skills and musical achievements[8]. The research by Phineas Magwati and Philemon Manatsa demonstrated that visually impaired learners could enhance auditory cognitive abilities and motor sensory skills through music while developing creativity and improvisational skills[9]. This paper confirms that musical activities allow visually impaired individuals to build self-efficacy, experience social recognition, and achieve personal fulfillment. Therefore, it proposes the development of a system that enables visually impaired individuals to perform more freely, fostering self-efficacy and a sense of accomplishment.

3. Drum Performance Analysis System Diagram

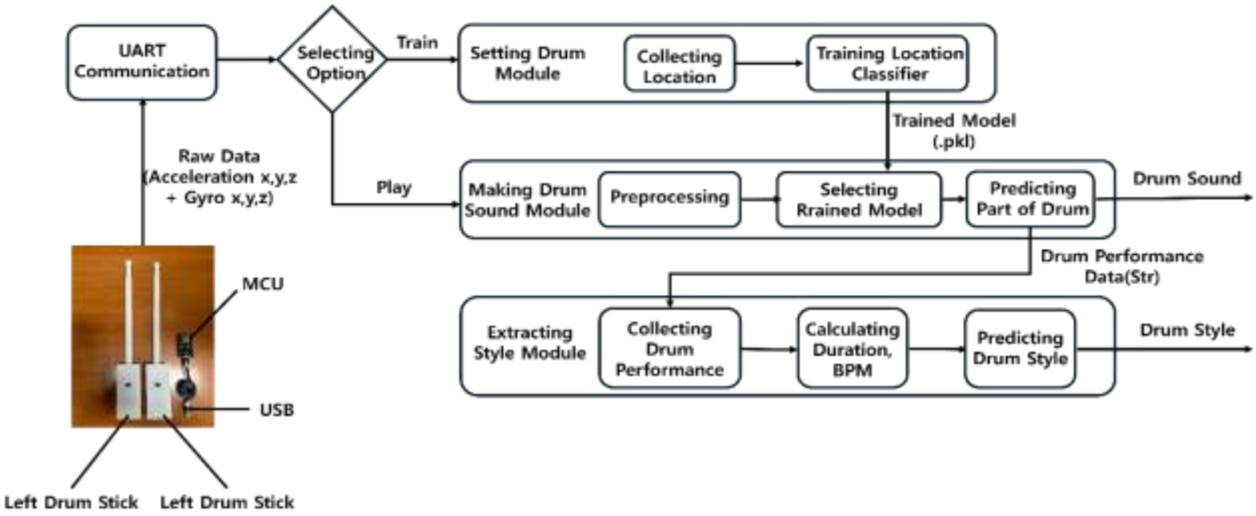

Fig. 1: Drum Performance Analysis System Diagram

The proposed system comprises three main modules: the Setting Drum Module, the Making Drum Sound Module, and the Extracting Style Module. When a visually impaired person uses a dedicated drumstick, the MCU of the drumstick inputs raw data consisting of acceleration x, y, z coordinates and gyro x, y, z coordinates to the user's PC via UART communication. Upon receiving the raw data, the user can select an action option using a push button. Suppose the visually impaired individual selects the Training option. They can collect raw data by playing their desired drum parts at their preferred locations using the dedicated drumstick according to the system's guidance[17]. This process collects performance data for a total of eight drum types: Crash Cymbal, Small Tom, Middle Tom, Ride Cymbal, Hi-Hat, Snare, Kick, and Floor Tom.

The Setting Drum Module utilizes the acceleration x, y, z coordinates, and gyro x, y, z values from the collected raw data of the eight drums to train classifiers using KNN and Random Forest algorithms, predicting the virtual positions of the eight drums. The Setting Drum Module then transfers the trained virtual drum position classifier to the Making Drum Sound Module as a pkl file, a serialized file format used in Python for saving machine learning models. When the Play option is selected, the Making Drum Sound Module processes the sensor values from the dedicated drumstick using StandardScaler, and the trained model from the Setting Drum Module is applied to predict which drum part is being played.

The Extracting Style Module collects the location and frequency of the drums being played and calculates the collection time. It gathers data on the time (seconds) during which the visually impaired individual is playing and the number of times each drum is struck. The total number of drum strikes is summed to calculate the BPM (Beats Per Minute) of the visually impaired individual's drumming. The time (in seconds) is then divided by 60 to convert into minutes. The BPM is calculated by dividing the total number of drum strikes by the time (minutes) during which the performance occurs.

The features for classification include the BPM of the visually impaired individual's drumming, the duration of play (in seconds), and the number of strikes for each drum type (Crash et al.), totaling ten features. These features are classified into styles such as Rock, Hip-Hop, Funk, and Punk using a One-vs-All DNN approach. This model learns binary classifiers that determine whether each class (4 drum styles) is present, ultimately identifying the class with the highest probability among the four options.

The predicted drum sound from the Making Drum Sound Module is played as an mp3 recording of the drum performance. The predicted drum style from the Extracting Style Module provides a string indicating which style (Rock et al., or Punk) the visually impaired individual played. The predicted drum style allows the visually impaired person to verify whether their performance aligns with their intentions, fostering a sense of accomplishment and self-efficacy through the process. Figure 1 shows the Drum Performance Analysis System Diagram.

4. Experiment

In our study, we utilized the Expanded Groove MIDI Dataset (E-GMD)[18], which contains comprehensive data on drum performances, to develop a model for predicting the drum style of visually impaired individuals within the Extraction Style Module. In this model, we defined style as the target variable and used BPM, Duration, Kick, Ride Cymbal, Hi-Hat, Small Tom, Middle Tom, Snare, Floor Tom, and Crash Cymbal as the feature values derived from the E-GMD[18]. BPM represents the number of strikes within a minute, while duration indicates the total time of performance measured in seconds. The features Kick, Ride Cymbal, Hi-Hat, Small Tom, Middle Tom, Snare, Floor Tom, and Crash Cymbal correspond to the count of strikes for each respective drum part. The Style category encompasses four primary classes: Rock, Hip-hop, Funk, and Punk. We ensured that all classes were utilized equally to address data imbalance issues. For the analysis of the user's drum style, we adopted several machine learning and deep learning approaches, including Ensemble Voting[19], Random Forest[20], Deep Neural Networks (DNN)[21], Convolutional Neural Networks (CNN)[22], Long Short-Term Memory (LSTM)[23], Gated Recurrent Units (GRU)[24], Ensemble Voting with DNN, Ensemble Bagging with DNN, and One-vs-All (OvA) with DNN. We trained the model on the MIDI dataset to compare and analyze the performance of these various approaches.

Table 1: Analysis of the preformance of the ensemble -based machine learning algorithms

Classifier	Hyper Parameter	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)
RandomForest	max_depth = 4, max_feature = 20, min_samples_leaf = 2, n_estimators = 400	97.56	57.50	79.87
Ensemble Voting	**Estimators = SVC, DecisionTree, KNN, voting = 'soft'**	**99.80**	**88.75**	**93.25**

To classify users' drum performance styles into four categories: Rock, Hip-hop, Punk, and Funk, we utilized various machine learning approaches, including Ensemble Voting, Random Forest, Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Ensemble Voting with DNN, Ensemble Bagging with DNN, and One-vs-All (OvA) with DNN. We conducted a comparative analysis of the performance of Random Forest, which is an ensemble method based on bagging, and Ensemble Voting. We configured the estimator as a Decision Tree to train the Random Forest model, with max_depth set to 4, max_features to 20, and min_samples_leaf to 2, while n_estimators was set to 400. The training results revealed that the Random Forest model achieved a Train Accuracy of 97.56%, Validation Accuracy of 57.50%, and Test Accuracy of 79.87%. In contrast, the Ensemble Voting model was established with estimators, including a Support Vector Classifier (SVC), Decision Tree, and K-Nearest Neighbors (KNN), and training was conducted using soft voting. The results from this model indicated a Train Accuracy of 99.80%, Validation Accuracy of 88.75%, and Test Accuracy of 93.25%. Ensemble Voting demonstrated superior performance among the ensemble methods utilizing machine learning algorithms. Table 1 presents an analysis of the performance of the ensemble-based machine learning algorithms.

Table 2: Analysis of the preformance of the deep learning models

Classifier	Hyper Parameter	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)
DNN	Epoch=500, Batch_size=32, 1 Dense Layer	96.29	97.34	**96.76**
CNN	Epoch=500, Batch_size=32, 1 Conv1D Layer	**99.37**	**99.00**	91.37
LSTM	Epoch=500, Batch_size=32, 1 LSTM Layer	92.66	96.13	96.25
GRU	Epoch=500, Batch_size=32, 1 GRU Layer	90.72	96.00	94.63

To classify drum performance styles, we employed deep learning techniques, specifically Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU). For the DNN model, we configured the architecture with a single Dense Layer as the hidden layer, which contained 256 hidden units and a total of 3,844 parameters. The training results indicated a Train Accuracy of 96.29%, Validation Accuracy of 97.34%, and Test Accuracy of 96.00%. In the CNN model, we utilized the Conv1D layer provided by Keras, again implementing a single hidden layer. This model included a total of 33,988 parameters. The training outcomes revealed a Train Accuracy of 84.83%, Validation Accuracy of 92.87%, and Test Accuracy of 90.55%. The LSTM model was structured with a single LSTM layer as the hidden layer, comprising 256 hidden units and a total of 19,460 parameters. The training results showed a Train Accuracy of 92.66%, Validation Accuracy of 96.13%, and Test Accuracy of 96.25%. For the GRU model, we similarly configured a single GRU layer as the hidden layer, consisting of 256 hidden units and a total of 14,852 parameters. The training outcomes indicated a Train Accuracy of 90.72%, Validation Accuracy of 96.00%, and Test Accuracy of 94.63%. Table 2 presents an analysis of the performance of the deep learning models employed in this study.

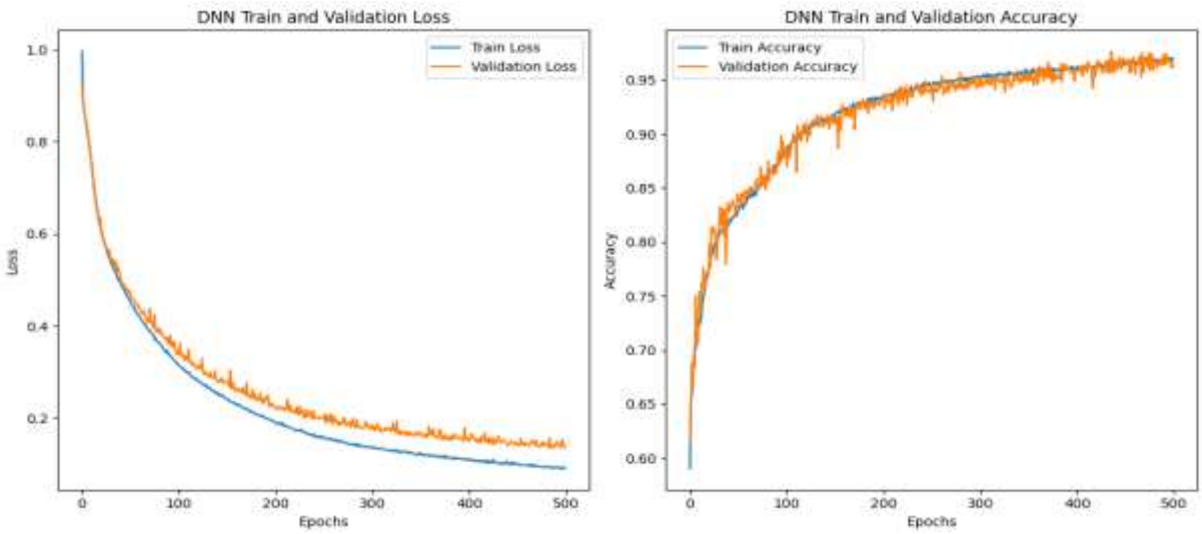

Fig. 2: Loss and Accuracy Graph for Classifying Drum Performance Style using DNN

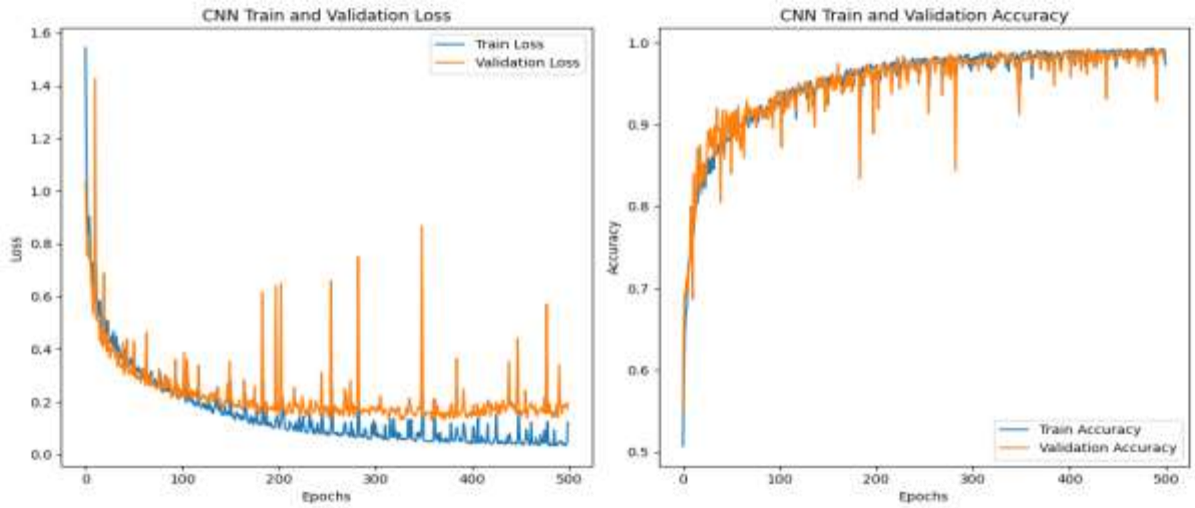

Fig. 3 Loss and Accuracy Graph for Classifying Drum Performance Style using CNN

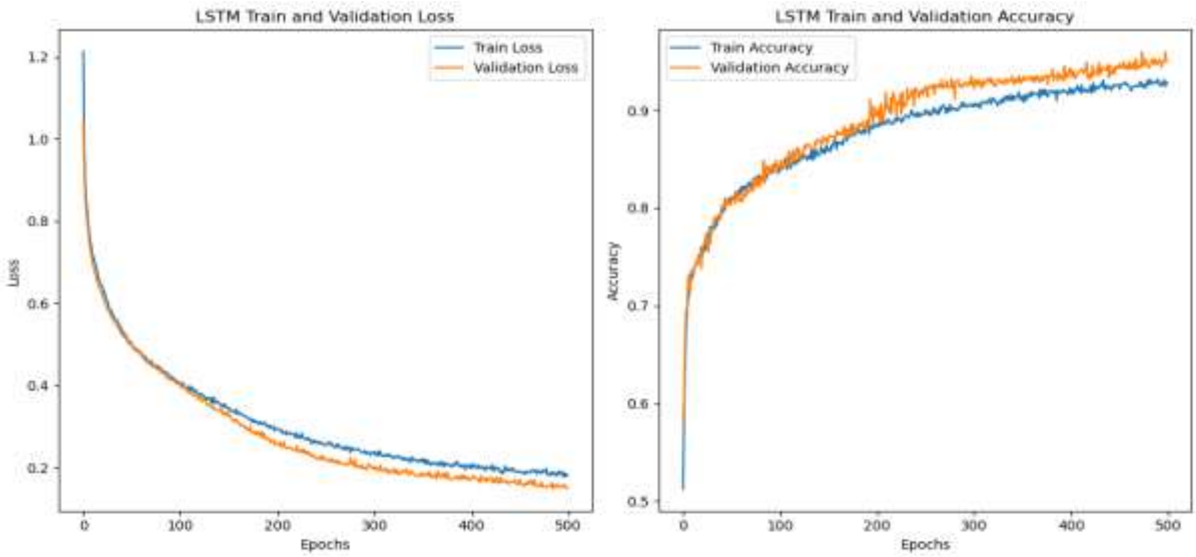

Fig. 4: Loss and Accuracy Graph for Classifying Drum Performance Style using LSTM

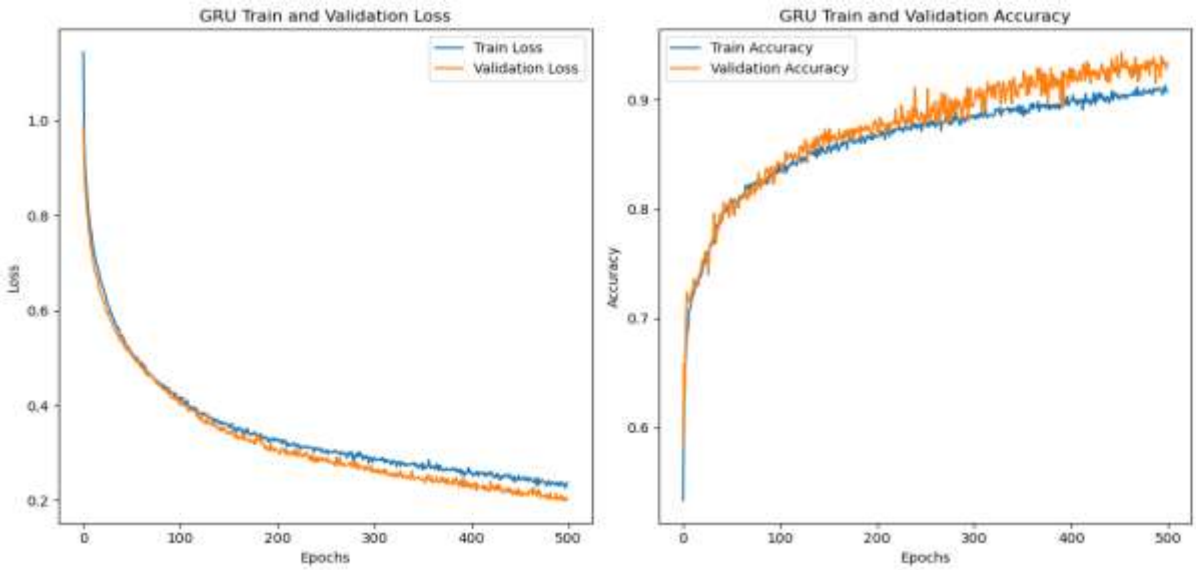

Fig. 5: Loss and Accuracy Graph for Classifying Drum Performance Style using GRU

The CNN model exhibited the highest performance in terms of Train Accuracy and Validation Accuracy; however, the DNN model achieved the highest Test Accuracy. In Fig. 3, the Validation Accuracy of the CNN model exhibited instability, with fluctuations and an increase in Validation

Loss after 200 epochs. CNN Loss and Accuracy Graph shows the presence of overfitting starting from 200 epochs. Therefore, this study utilized the DNN model to predict the drum performance style of visually impaired individuals, as it demonstrated no signs of overfitting and provided the highest performance in Test Accuracy. Figures 2, 3, 4, and 5 present the graphs showing the Train Loss, Validation Loss, Train Accuracy, and Validation Accuracy for each of the models: DNN, CNN, LSTM, and GRU.

Table 3: Performance of Four DNN Models

Classifier	Hyper Parameter	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)
DNN_1	1 Dense Layer , Hidden Unit 256	96.29	97.34	96.76
DNN_2	2 Dense Layer , Hidden Unit 256, 128	96.15	98.75	96.57
DNN_3	3 Dense Layer, Hidden Unit 256, 128, 64	**97.17**	**99.30**	**98.34**
DNN_4	4 Dense Layer, Hidden Unit 256, 128, 64, 32	96.31	98.05	97.13

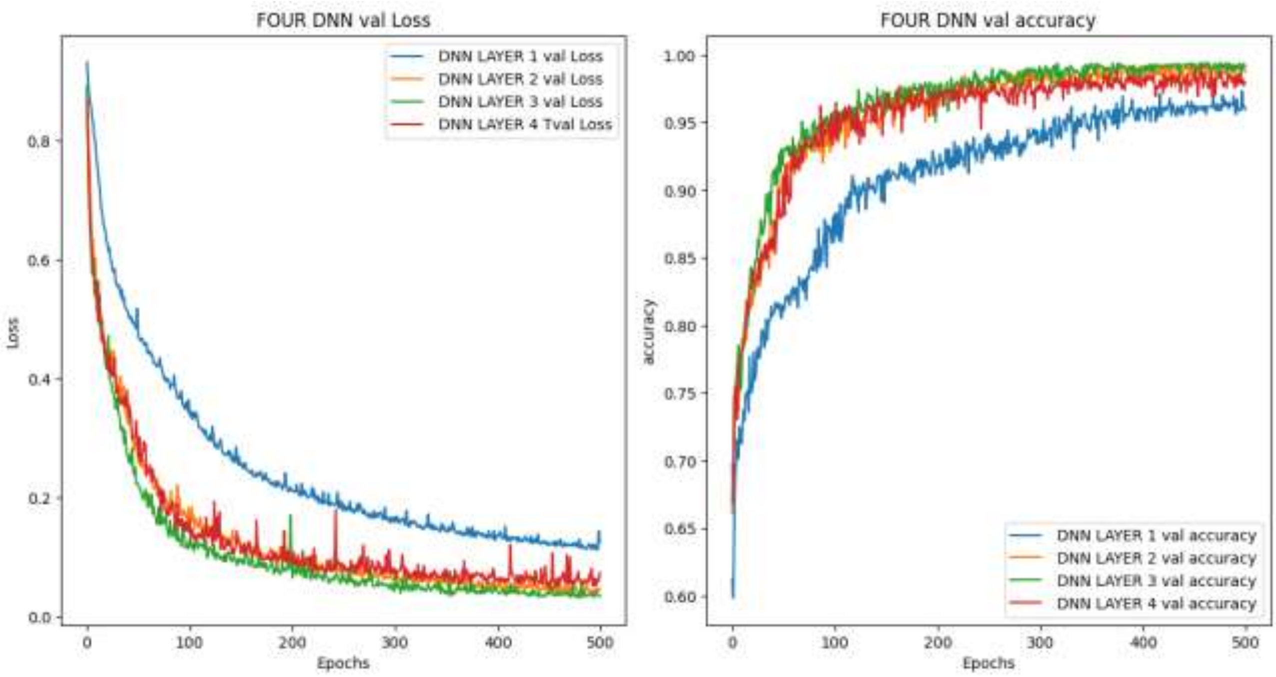

Fig. 6: Graph of Four DNN Models

To identify the most performant model among the DNN-based architectures, we conducted a comparative analysis of four models configured with 1, 2, 3, and 4 layers, respectively.

- DNN_1 consists of a single Dense Layer with 256 Hidden Units, utilizing a total of 3,844 parameters for training.
- DNN_2 comprises two dense layers, with the first layer set to 256 hidden units and the second layer set to 128 hidden units. It employs a total of 36,228 parameters for training.
- DNN_3 incorporates three dense layers, with 256 hidden units in the first layer, 128 in the second layer, and 64 in the third layer. It utilizes 44,228 parameters for training.
- DNN_4 features four Dense Layers, structured with 256 Hidden Units in the first layer, 128 Hidden Units in the second layer, 64 Hidden Units in the third layer, and 32 Hidden Units in the fourth layer, employing a total of 46,180 parameters for training.

The performance analysis revealed that DNN_3 achieved the highest accuracy, with a Train Accuracy of 97.17%, a Validation Accuracy of 99.30%, and a Test Accuracy of 98.34%.

Consequently, we adopted DNN_3, configured with three layers, for our study. Table 3 provides a detailed analysis of the DNN models' performance in this research. Table 3 presents performance of four DNN models. Figure 6 presents graph of four DNN models

Classifier	Hyper Parameter	Test Accuracy (%)
Ensemble Voting with DNN	estimators = 5 DNN(epoch 100, batch_size 32)	95.62
Ensemble Bagging with DNN	estimators = 5 DNN(epoch 100, batch_size 32)	97.12
OvA with DNN	**estimators = 4 DNN(epoch 100, batch_size 32)**	**99.95**

Table 4: Comparison between the ensemble-based and the One-vs-All (OvA)-based DNN.

We utilized the DNN model, which exhibited the highest performance among our deep learning architectures, to implement ensemble techniques, specifically voting, bagging, and One-vs-All (OvA) methods. The ensemble voting model based on DNN was configured with five DNN models, utilizing soft voting with parameters set to epoch = 100 and batch_size = 32. The training results yielded a Test Accuracy of 95.62%. To train the bagging model utilizing DNN, we configured five DNN models, each trained on distinct data samples with the same parameters (epoch = 100, batch_size = 32). This model achieved a Test Accuracy of 97.12%. We trained four classifiers for the OvA model with DNN, each corresponding to one of the four classes (Rock et al.), verifying whether the input belonged to the respective class. These classifiers were based on the DNN model with parameters set to epoch = 100 and batch_size = 32. The training results for the OvA model revealed a remarkable Test Accuracy of 99.95%. Comparative analysis between the DNN-based ensemble techniques and the OvA method demonstrated that the OvA technique with DNN exhibited the highest performance. Table 3 presents a comprehensive analysis of the performance of the ensemble techniques and OvA methods utilizing DNN. In total, we trained and analyzed nine classifiers to assess the drumming style of visually impaired individuals: Random Forest, Ensemble Voting, DNN, CNN, LSTM, GRU, Ensemble Voting with DNN, Ensemble Bagging with DNN, and One-vs-All with DNN. Given that the One-vs-All with DNN achieved the highest Test Accuracy of 99.95%, we adopted this method to classify the drumming styles of visually impaired individuals. Table 4 presents the performance cComparison between the ensemble-based and the One-vs-All (OvA)-based DNN.

5. Conclusions

This study implemented a system that analyzes the drumming performance of visually impaired individuals and provides insights into their playing styles. The system collects drumming data through dedicated drumsticks and analyzes the performance style, enabling users to confirm whether the system accurately interprets their intentions. This interaction fosters a sense of self-efficacy among visually impaired individuals, allowing them to feel empowered in their physical movements and express emotions through music, thereby enhancing their mental well-being. Future research will focus on utilizing the analyzed drumming styles of visually impaired individuals to recommend performance pieces within the same genre and generate corresponding sheet music.

Acknowledgment

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2022S1A5C2A07091326)

References

[1] Ruxandra Tapu, undefined. Bogdan Mocanu, undefined. Titus Zaharia. "Wearable assistive devices for visually impaired: A state of the art survey," in Pattern Recognition Letters, vol. 137, pp. 37-52, 2020.

[2] Lavric, Alexandru et al. "A Comprehensive Survey on Emerging Assistive Technologies for Visually Impaired Persons: Lighting the Path with Visible Light Communications and Artificial Intelligence Innovations." Sensors (Basel, Switzerland) vol. 24,15 4834. 25 Jul. 2024

[3] Yen, C., et al. "Association of Visual Impairment with Psychological Distress in Older Adults: A Survey of 105,092 Older People in Taiwan," in Journal of Clinical Medicine, vol. 11, pp. 1458, 2022.

[4] Fredéric Gougoux, Franco Lepore, Maryse Lassonde, Patrice Voss, Robert J Zatorre, Pascal Belin. Neuropsychology: Pitch discrimination in the early blind. Nature 2004,430(6997):309.

[5] Magwati, Phineas, Philemon, Manatsa. "The musical development of visually impaired learners: A case study of a special needs secondary school". African Musicology Online 12. (2023): 1-10.

[6] Kern P, Wolery M. "Participation of a preschooler with visual impairments on the playground: effects of musical adaptations and staff development," J Music Ther. 2001, 38(2):149-64.

[7] Faby K. "Music Therapy for Children with Cerebral Visual Impairment," Florida State University, 2016.

[8] Dauber, Markus. "What is Essential is Invisible to the Eye" - A Music Therapy Tale of a Young Woman with Visual Impairment." 2011.

[9] Magwati , P. ., and P. . Manatsa. "The Musical Development of Visually Impaired Learners: A Case Study of a Special Needs Secondary School". African Musicology Online, vol. 12, no. 1, May 2023, pp. 1-10

[10] David Baker, Lucy Santos Green. "Perceptions of schooling, pedagogy and notation in the lives of visually-impaired musicians." Research Studies in Music Education 2016, vol.38, pp.193 - 219.

[11] Osborne, Lucy. An Exploration of Music Therapy Practice with Children with Visual Impairment and Profound and Multiple Learning Disabilities (PMLD). 2019.

[12] L. Turchet, D. Baker, T. Stockman, "Musical Haptic Wearables for Synchronisation of Visually-impaired Performers: a Co-design Approach," in Proceedings of the 2021 ACM International Conference on Interactive Media Experiences, 2021, pp. 20–27.

[13] D. Zaveri, R. Parekh, R. Lobo, S. Ardekar, V. Vora and A. Kore, "Aero Drums - Augmented Virtual Drums," 2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), Goa, India, 2022, pp. 516-520.

[14] Lu Leon, Kang Jin, Crispin Chase, Girouard Audrey. :Playing with Feeling: Exploring Vibrotactile Feedback and Aesthetic Experiences for Developing Haptic Wearables for Blind and Low Vision Music Learning," 2023.

[15] B. Yuan, E. Folmer, "Blind hero: enabling guitar hero for the visually impaired," in Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, 2008, pp. 169–176.

[16] Magwati , P. ., and P. . Manatsa. "The Musical Development of Visually Impaired Learners: A Case Study of a Special Needs Secondary School". African Musicology Online, vol. 12, no. 1, May 2023, pp. 1-10

[17] Jonguk Jung and Yoosoo Oh. "Customized Drum System for Visually Impaired People Based on Machine Learning Classification Algorithm." 한국 HCI 학회 학술대회, 2024, pp. 11-17

[18] Lee Callender, Curtis Hawthorne and Jesse Engel, "Improving Perceptual Quality of Drum Transcription with the Expanded Groove MIDI Dataset," 2020.

[19] Akinsola, J E T. "Supervised Machine Learning Algorithms: Classification and Comparison." International Journal of Computer Trends and Technology (IJCTT),2017, pp.128 - 138.

[20] T. Dietterich, "Ensemble Methods in Machine Learning," in Multiple Classifier Systems, 2000, pp. 1–15.

[21] Güvenç, E, Çetin, G, Koçak, H. "Comparison of KNN and DNN Classifiers Performance in Predicting Mobile Phone Price Ranges." Advances in Artificial Intelligence Research 2021, vol.1, no.1, pp.19–28.

[22] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems , 2012, 25.

[23] Staudemeyer, Ralf C., and Eric Rothstein Morris. "Understanding LSTM--a tutorial into long short-term memory recurrent neural networks." arXiv preprint arXiv:1909.09586, 2019.

[24] FU, Rui; ZHANG, Zuo; LI, Li. Using LSTM and GRU neural network methods for traffic flow prediction. In: 2016 31st Youth academic annual conference of Chinese association of automation (YAC). IEEE, 2016. p. 324-328.