ISSN: 2508-7894
KJAI website: https://accesson.kr/kjai
doi: http://doi.org/10.24225/kjai.2024.12.3.33

A Study on the Performance Improvement of MLP Model for Kodály Hand Sign Scale Recognition

Na Gyeom YANG[1], Dong Kun CHUNG[2]

Received: April 14, 2024. Revised: May 10, 2024. Accepted: September 05, 2024.

Abstract

In this paper, we explore the application of Kodály hand signs in enhancing children's music education, performances, and auditory assistance technologies. This research focuses on improving the recognition rate of Multilayer Perceptron (MLP) models in identifying Kodály hand sign scales through the integration of Artificial Neural Networks (ANN). We developed an enhanced MLP model by augmenting it with additional parameters and optimizing the number of hidden layers, aiming to substantially increase the model's accuracy and efficiency. The augmented model demonstrated a significant improvement in recognizing complex hand sign sequences, achieving a higher accuracy compared to previous methods. These advancements suggest that our approach can greatly benefit music education and the development of auditory assistance technologies by providing more reliable and precise recognition of Kodály hand signs. This study confirms the potential of parameter augmentation and hidden layers optimization in refining the capabilities of neural network models for practical applications.

Keywords: Kodály Hand Sign, Artificial Neural Networks (ANN), Multilayer Perceptron (MLP), Parameter Augmentation, Hidden Layers Optimization

Major Classification Code: Artificial Intelligence

1. Introduction

1.1. Research Background

The Kodály hand signs (hand symbols) represent a means of expressing music for the hearing impaired, as well as a widely utilized tool for the musical education of children. These hand signs are effective in developing a sense of pitch as well as rhythm, due to their ability to distinguish musical tones. Consequently, Kodály hand signs are recognized as a vital tool in music classes, fostering active participation and aural development in students. The commonly used Kodály hand signs scale has been revised and supplemented by the Hungarian composer and music educator Zoltán Kodály, including semitones such as "Pa" and "Ti," enhancing its functional use in music education (Yun et al., 2021).

With the rapid societal changes and technological advancements, methods in music education are becoming increasingly sophisticated. This study aims to focus on the

1 First Author. Undergraduate Student, Medical IT, Eulji University, Korea. Email: hanbiiii@naver.com
2 Second Author, Professor, Medical IT, Eulji University, Korea, Email: tchung@eulji.ac.kr

improvement of children's music education through the use of Kodály hand signs. Ultimately, the goal is to enhance the recognition rate of the Kodály hand signs scale by employing MLP models and artificial neural network techniques.

Neural networks have been proven effective in pattern classification. As they can recognize patterns from existing data and learn from them, they can effectively classify user-input patterns once the data is trained, offering convenience in use. Due to their pattern recognition capabilities, they adapt quickly to anomalies, even if sensor values fluctuate unexpectedly, and they require the input of patterns rather than identical values (Jang et al., 2012).

Contact-based motion recognition methods that rely on accelerometers or gyro sensors offer the advantage of directly measuring user movements, unlike camera-based visual information. However, these methods face challenges, such as the need to segment continuous motion patterns into intended motion command units and to handle uncertain information like unintended movements since they recognize motions solely based on acceleration and angular velocity values in three-dimensional space (Jang et al., 2012).

Previous studies have identified difficulties in recognizing hand sign scales, particularly in distinguishing the values between continuous scales such as "Do" to "Re." Such errors have been shown to negatively impact the overall recognition rate of scales.

This study intends to effectively process the intermediate values of continuous scales through optimization techniques derived from model training.

The significance of this study lies in minimizing the errors that occur during the recognition of intermediate values of specific scales. By preprocessing the results of intermediate scale recognition to reduce factors that hinder recognition rates and adjusting the number of hidden layers through the application of the MLP model, this study aims to propose ways to optimize the use of Kodály hand signs in children's music learning.

1.2. Research Purpose

This study contributes to enhancing the recognition rate of the Kodály hand sign scale, widely applied in music education for individuals with hearing impairments and children. By investigating methods to optimize a Multilayer Perceptron (MLP) model for Kodály hand sign scale recognition, the focus was on increasing the recognition rate. Particularly, by enhancing the model's ability to capture subtle variations between consecutive scales, it supports music educators and learners in achieving accurate musical performance. This study is expected to enhance the recognition rate of the MLP model through new technological improvements, thereby enhancing the accessibility and efficiency of music education and contributing to the improvement of musical expression skills.

1.3. Previous Research

In the field of dynamic hand gesture recognition, the CRNN (Convolutional Recurrent Neural Network) model has been utilized effectively for classifying complex patterns of both static and dynamic hand gestures, demonstrating robust capabilities. In relevance to this paper, Kwon Yong-seong's study on the CRNN architecture contributes significantly in terms of the universality of training data using hand skeleton images (Kwon, 2023). This provides a research direction for accurate recognition of Kodály hand sign scales, and this study explores the possibility of achieving precise recognition rates for various hand signs by applying the CRNN model. The experimentally validated CRNN model structure and the learning method using hand skeleton images provide important evidence for accurately understanding the characteristics of complex gestures aimed at recognizing Kodály hand signs in this study. Additionally, hand gesture recognition, which plays a crucial role in efficient human-machine interfaces (HMI), has shown great potential in accurately capturing subtle movements of hand signs using wearable sensor technology, as demonstrated in Lee Hyun-kyu's research (Lee, 2022). The system implemented in this study achieved a high recognition accuracy of 98.7% on low-power embedded devices, providing a reliable benchmark for recognizing Kodály hand sign scales.

Lee Soo-jin and Kang Ji-heon significantly improved the accuracy of hand gesture recognition using a novel lightweight deep learning model based on point cloud data from mmWave radar (Lee et al., 2023). The research team adopted a 2D projection method for preprocessing input data, developing a 2D-CNN-TCN model that achieved a high recognition rate of 95.06% while drastically reducing the number of learning parameters. This model demonstrated superiority in recognition accuracy compared to the traditional 3D-CNN-LSTM model, suggesting an important direction for achieving the goal of enhancing "the recognition rate of Kodály hand sign scales." By showcasing the practical applicability of lightweight deep learning models, this research provides important insights into real-time processing capabilities and model lightweighting pursued within its scope.

According to the study by Bae Seong-jun and Kim Hyeong-seok (Bae et al., 2018), it has been demonstrated that deep learning algorithms utilizing IMU

(Inertial Measurement Unit) sensors can improve gesture recognition rates. Although these studies improved recognition accuracy and error update speed by applying specific activation functions, they had limitations in fully reflecting the complex characteristics of hand gestures, which could lead to the recognition of unnecessary movements. To overcome these limitations, this study aims to enhance the recognition rate of MLP models, focusing on the movements transitioning from “Do” to “Re” and “Re” to “Mi” in the scale, which are crucial for recognizing Kodály hand signs.

Ha Sim-hyeong, Lim Seong-bin, and Choi Woo-kyeong explored the applicability of neural networks and backpropagation algorithms for classifying motionpatterns by continuously analyzing sensor data in previous studies (Ha et al., 2006). While they achieved high recognition rates for simple gestures using a system combining various sensors to recognize hand gestures, they found limitations in recognizing complex gestures. Pedro and L. Jorgerk improved gesture recognition using 3-axis accelerometer sensors for dynamic motion (T.Pedro et al., 2011) and Jang Min-seon, Choi Hyeon-ho, Kim Ji-hwan, and Lee Seong -il enhanced recognition rates by combining accelerometer and gyroscope sensor data with decision trees and neural networks (Jang et al, 2012).

Although studies on user motion recognition using accelerometers have been widely conducted, existing methods have shown limitations in recognizing unintentional movements by users.

Based on the limitations of previous studies mentioned above, this study aims to develop an artificial neural network model capable of recognizing complex movements of Kodály hand sign scales and considering unintentional movements by users. Specifically, it focuses on optimizing the structure of MLP models, increasing the amount of training data to improve recognition rates for various complex movements, and emphasize enhancing children’s musical development and the quality of classes.

2. Main Body

2.1. Research Method

This study was conducted in three main stages: data collection, data preprocessing, and design of the artificial neural network MLP model.

2.1.1. Data Collection

Firstly, to develop a music performance model, equipment capable of recognizing hand gestures for Kodály hand sign scales was established. For precise motion recognition of hand gestures, ESPRESSIF’s ESP 8266-12E model was utilized. A board was assembled by combining an MPU6050 gyro sensor capable of measuring hand gesture acceleration and angular velocity with a bending sensor capable of measuring the degree of hand bending. Through this setup, we digitized hand gesture recognition and collected data. Through this setup, hand gesture recognition was digitized, and data was collected.

The hand gesture recognition glove was connected to Arduino to confirm data input and output, and the collected data was processed and preprocessed using the program “Cool Term.” After setting up the equipment, the sensor port was connected to the hand gesture recognition glove to collect hand gesture data for Kodály hand sign scales “Do,” “Re,” “Mi,” “Fa,” “Sol,” “La,” and “Ti.” The measured values were stored in CSV files for utilization. Figure 3 illustrates the Kodály hand sign scales, and Table 1 represents the hand gestures used during actual data collection.

Table 1: A list of collected hand gestures

	
Do	Re
	
Mi	Fa
	
So	La
	
Ti	Do

2.1.2. Data Preprocessing

Data preprocessing is one of the key steps in enhancing the learning and prediction performance of the MLP model.

(1) Preprocessing in Data Collection

Scaling is a data preprocessing task that facilitates the comparison of data by making the range of data appear similar. The raw acceleration and angular velocity data input from the MPU6050 sensor are transformed through data scaling to be represented in radians, and the values of variables in model training can be adjusted to a specific range of [0, 1] or [-1, 1] through the normalization process.

(2) Preprocessing in Csv Files - Classification

To accurately train the model with large volumes of data stored in csv files, a classification process of the data must be preceded. That is, the data must be classified by musical scale. Through this, the model can learn and differentiate movements corresponding to each musical scale, and it becomes capable of predicting and recognizing movements related to a specific scale. Moreover, it helps the model to understand and learn the musical meaning, and if the data is differentiated reflecting the relationships between musical scales, the learning and classification of the model can be conducted more efficiently.

In this study, integer values from 1 to 7 were assigned to the musical scales from “Do” to “Ti,” and a separate number “0” was given to unspecified intermediate values to classify them as “scales without sound.” This procedure is a widely proven method in previous studies related to music recognition and music performance, known as an effective way to improve the model’s performance by distinguishing data into musical scales and intermediate value

2.2. Artificial Neural Network MLP Model Design

In this study, the Backpropagation Algorithm was utilized to develop and analyze a music performance model using the MPU6050 gyro sensor. The backpropagation algorithm adjusts the weights and biases of each layer to learn the relationship between input data and the desired output, and it was used to train and predict the neural network of this study’s model. Figure 1 illustrates the overall stages of the model algorithm, and Table 2 presents the algorithm in pseudocode.

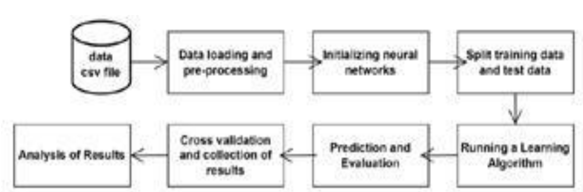

Figure 1: Model algorithm flow chart

This model represents a Multi-Layer Perceptron (MLP) form composed of hidden layers and an output layer. The hidden layer plays a role in learning the nonlinear characteristics of the data, and in this study, the hidden layer is configured as a single dense layer. The hyperparameter, which is the recognition rate, plays an important role in accelerating the model’s convergence and maintaining stability. For this reason, the Epochs value was set to “500” in this study, ensuring a sufficient number of learning iterations for the model to adequately learn the data patterns. Moreover, the cross-validation method was utilized to divide the data into several folds. Each fold alternately performs the role of training and testing data, assessing the model’s generalization performance in a reliable manner. In this study, the k-fold cross-validation procedure was employed, with the k value set to 5.

The algorithm (Table 2) presented here outlines the procedure for applying the backpropagation training method using Stochastic Gradient Descent on the Seeds Dataset. It details the steps from loading and preprocessing the dataset to training the neural network and evaluating its performance. Specifically, the algorithm converts string data to numeric formats, normalizes the data, and divides the dataset into k folds for cross-validation. The neural network is trained with specified learning parameters (learning rate, number of epochs, number of hidden layers) and is evaluated on its ability to accurately predict outcomes. The final output is the set of accuracy scores from the cross-validation process.

Table 2: Algorithm pseudocode

Algorithm 1 Backpropagation on the Seeds Dataset

Require: Function to load dataset, Function to convert string columns to floats, Function to convert string columns to integers, Function to find min and max values for each column, Function to normalize dataset to 0-1 range, Function to split dataset into k folds, Function to calculate accuracy metric, Function to calculate neuron activation for inputs, Function to pass neuron activation through transfer function, Function to forward propagate inputs through the network to get outputs, Function to calculate derivative of neuron output, Function to propagate errors backward and store in neurons, Function to update network weights using errors, Function to train network for fixed number of epochs, Function to initialize neural network, Function to make predictions using the network, Backpropagation algorithm using Stochastic Gradient Descent

procedure BACKPROPAGATION(*filename*, n_f*olds*, l_r*ate*, n_e*poch*, n_h*idden*)
  Load and prepare the dataset
  dataset ← LOADCSV(*filename*)
  for i ← 0 **to** *len*(*dataset*[0]) − 1 **do**
    STRCOLUMNTOFLOAT (*dataset*, *i*)
  end for
  STRCOLUMNTOINT(*dataset*,*len*(*dataset*[0])− 1)

```
    minmax ← DatasetMinMax(dataset)
    NormalizeDataset(dataset, minmax)
    Evaluate the algorithm
    scores ← EvaluateAlgorithm(dataset, BackPropagation, n_folds, l_rate, n_epoch, n_hidden)
    return scores
end procedure
```

This section is not a table but a detailed procedural description of the backpropagation algorithm as applied in our study. It provides a structured sequence of operations, from data preparation to the training and evaluation of the model, ensuring clarity on the methodology used to achieve the results reported in the study.

2.3. Data Learning Process

For data learning, the collected data was first attempted for learning after excluding the intermediate values. That is, the data from "Do" to Ti" were used discontinuously, and the learning results showed a recognition rate of 86.579%. Next, including the intermediate values and training the

The following insights were derived from the experimental results. The total dataset used in the experiment consisted of 9,780 movements, and the MLP model utilizing additional angular velocity data achieved a recognition rate of 86.226% (refer to Figure 8).

This indicates that the model performance was improved with the addition of parameters. However, the degree of performance improvement was not significant. This result suggests that the impact of the added parameters on model learning was minimal.entire data set resulted in a recognition rate of 79.463%. This indicates that including intermediate values can decrease the data recognition rate. Therefore, this study considers increasing the number of parameters and hidden layers to improve the decreased recognition rate.

2.3.1. Adding Parameters

Adding parameters is one of the important methods to improve the learning and prediction performance of machine learning and MLP models, generally used to adjust the complexity of the model. This study aims to add parameters to improve the performance of the model, which has been decreased by including intermediate value data. The new parameters need to be designed to more accurately capture the spatial characteristics of hand movements. Therefore, the characteristics of angular velocity were used to add angular velocity data itself as a parameter.

2.3.2 Increasing the Number of Hidden Layers

Another method to improve the expressiveness of the model is to increase the number of hidden layers. By increasing the number of hidden layers, the model forms a multi-layer neural network that can more delicately represent complex data patterns, making it applicable to this study's data, which includes a wide range of intermediate values and numerous hand movements.

Typically, the number of hidden layers and the performance of the model are proportional, but too many hidden layers can lead to overfitting problems. Thus, choosing the appropriate number of hidden layers is of utmost importance. In this study, the initial model was set with five hidden layers, and to compare the performance of the model according to the number of hidden layers, the layers were set to 6, 7, 8, and their results were compared.

2.4. Experimental Results

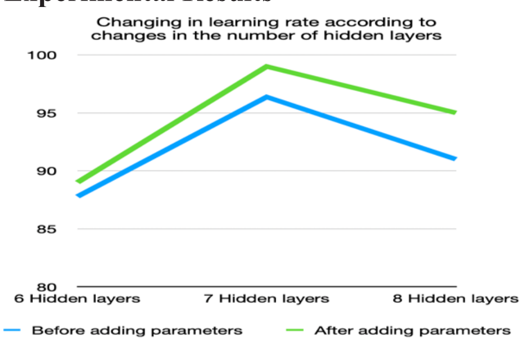

Figure 2: Changes in learning rate according to change in the number of hidden layer

This indicates that the model performance was improved with the addition of parameters. However, the degree of performance improvement was not significant.

This result suggests that the impact of the added parameters on model learning was minimal.
Furthermore, experiments that increased the number of hidden layers without considering parameters showed recognition rates of 87.783%, 96.370%, and 91.401% for 6, 7, and 8 hidden layers, respectively. Experiments that added angular velocity parameters in addition to increasing the number of hidden layers resulted in recognition rates of 88.602%, 98.927%, and 95.016%, respectively. According to these results, the highest recognition rate was observed with 7 hidden layers, and the addition of angular velocity parameters further improved the recognition rate. This indicates that the added hidden layers effectively reflected the model's complexity, allowing it to better perceive the sophisticated variations in the data.

However, the recognition rate decreased when the number of hidden layers was increased to 8, which could be attributed to overfitting. This phenomenon suggests excessive fitting to the training data, leading to a decreased ability to generalize to new data. This could limit the usability of the model in the actual music performance motion recognition process, necessitating the consideration of overfitting prevention techniques such as dropout and early stopping.

These results imply that there is a need to determine the optimal conditions for hidden layers and parameters to improve the accuracy of movement recognition using the MLP model.

3. Conclusion

3.1. Interpretation of Results

Through the analysis of experimental results, it was observed that the change in recognition rate obtained by varying the number of hidden layers was also different before and after the addition of parameters.

As can be seen from the results above, for the model without the addition of angular velocity parameters, the highest recognition rate was observed with 7 hidden layers, and increasing the number of hidden layers to 8 resulted in overfitting, as indicated by a decrease in recognition rate.

This suggests that additional hidden layers increase the complexity of the model, and beyond a certain level of complexity, the model’s ability to generalize to new data can be impaired. After considering the addition of angular velocity parameters, it was noted that the recognition rate improved further. These results demonstrate that angular velocity data effectively improves the model’s learning capability. In other words, adding new parameters and hidden layers to the neural network model is effective for enhancing the pattern recognition rate for a specific dataset, aligning with the outcomes of previous research. (Kim & Lee, 2022).

However, it is necessary to examine the overfitting phenomenon that occurs when increasing the number of hidden layers from 7 to 8. This phenomenon is commonly seen in complex neural network models and indicates that the model is overly optimized to the patterns in the training data, failing to reflect subtle variations in real phenomena. These findings underscore the importance of selecting an appropriate number of hidden layers, which is a criticalfactor in maintaining the generalization ability of the neural network model and improving the recognition rate of movements.

3.2. Summary of Results

This study explored methods to enhance the performance of an MLP-based neural network model capable of effectively recognizing music performance movements utilizing Kodály hand signs. Specifically, the addition of parameters and the adjustment of the number of hidden layers were used to maximize the model’s recognition rate for hand sign scales, with a focus on comparing the performance of recognition rates through the application of these methods.

By performing predictions on test data and comparing these outcomes to their actual values to measure accuracy, it was discovered that incorporating parameters using angular velocity data is an effective strategy for improving the model’s recognition rate, though its effect was found to be limited. Conversely, when the increase in the number of hidden layers was considered, a significant improvement in the model’s recognition rate and generalization ability was observed. These findings suggest that enhancing the model to internalize the complex characteristics of the data enables more precise recognition of music performance actions.

3.3. Implications

This study not only identifies optimal strategies for enhancing MLP model performance in recognizing Kodály hand sign scales but also opens avenues for practical applications in children's music education. By fine-tuning the number of hidden layers and integrating angular velocity data, we have demonstrated that these modifications substantially improve the model’s ability to interpret complex musical gestures accurately. These findings are pivotal for developing more responsive and effective music education tools that can adapt to the diverse needs of learners.

Furthermore, the enhanced recognition capabilities of our model could lead to more interactive and engaging music education experiences. For instance, educators could employ this technology to monitor and adjust teaching strategies in real-time, ensuring that all students are engaged and learning effectively. The potential for this technology to contribute to personalized learning environments represents a significant advance in educational methodologies.

Looking ahead, future research could explore the integration of additional sensory inputs, such as audio feedback, to further refine the model’s accuracy and responsiveness. Moreover, expanding the dataset to include a broader range of hand signs and environmental conditions would help in generalizing the model’s applicability to various learning contexts.

In conclusion, the enhanced understanding of MLP models in recognizing music performance actions not only advances our theoretical knowledge but also has significant implications for practical applications in music education. As we continue to explore these technologies, their integration into educational platforms is expected to enrich the learning experience and broaden the educational impact of music.

References

Bae, Seong-jun and Kim, Hyung-seok. (2018). A Study on the Improvement of Gesture Recognition Rate in IMU Sensor Systems Using Deep Learning. Proceedings of the Korean Institute of Communications and Information Sciences Conference, 304-305.18). *Proceedings of the Korean Institute of Communications and Information Sciences Conference*, 304-305.

Ha Shim-hyeong, Lim Syeong-bin, Choi Woo-kyung, Seo Jae-yong, Jeon Hong-tae. (2006). Development of a Motion Pattern Classification System Using Neural Networks. *Proceedings of the Korean Institute of Electrical Engineers Conference*, 897-898.

Jang, Min-seon, Choi, Hyun-ho, Kim, Ji-hwan, Lee, Sung-il. (2012). Development of a Hand Gesture Recognition System Using Neural Networks. *Proceedings of the Korea HCI Conference*, 676-681.

Kim Gwang – jin, Lee Chil-woo (2022). Time Series Prediction for Music Generation Using a Bi-LSTM Model. *Smart Media Journal*, *11*(10), 65-75.

Kim Jung-kyun, Lee Kang-bok, Hong, Sang-ki. (2016). Action Recognition Using Deep Neural Networks Based on IMU Sensors. *Proceedings of the Korean Institute of Information Scientists and Engineers*, 532-534.

Kwon Yong-sung. (2023). *Study on CRNN Structure for Dynamic Hand Gesture Recognition.* Master's Thesis, Kumoh National Institute of Technology, Graduate School. 49.

Lee Hyung-kyu. (2022). Hand Gesture Recognition Technology Based on Lightweight Artificial Neural Networks Using Wearable Sensors. *Journal of the Korean Society of Embedded Systems*, *17*(4), 229-237, DOI: 10.14372/IEMEK.2022.17.4.229

Lee Soo-jin and Kang Ji-heon. (2023). Lightweight Deep Learning Model for Hand Gesture Recognition Based on mmWave Radar Point Cloud. *Journal of the Institute of Control, Robotics and Systems*, *29*(9), 711-716, DOI: 10.5302/J.ICROS.2023.23.0096.

T.Pedro & L.Jorge. (2011). Distributed Accelerometers for Gesture Recognition and Visualization. *Advanced in information and Communication Technology*, Vol. 349/2011, 215-223.

Yun,Young Bae, Kwon,Young Man, Lim, Myung Jae, Park, Jeong Jun & Chung, Dong Kun. (2021). Development of Kodály Hand Sign Educational Tools using Multi-Layer Perceptron. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(13).