J. Electrical Systems 20-3s (2024): 2574-2587

[1]LingCao Wu*

Hierarchically Gated Recurrent Neural Network optimized with Tasmanian Devil Optimization for Piano Playing Based on Virtual Reality Technology

Abstract: - Playing the piano is a beautiful and extremely technical art form in addition to an elegant performance art. The player's ability to execute with elegance and technicality which requires years of practice will be crucial. The success of a piano concert ultimately rests not only on the performing abilities of the pianist but also on the pianist's mental and psychological makeup. This manuscript presents a Hierarchically Gated Recurrent Neural Network (HGRNN) optimized with the Tasmanian Devil Optimization (TDO) for predicting the real time instruction for playing piano based on virtual reality (PPVR-HGRNN-TDO). Initially, the data is collected from MIDI dataset. Afterward, the data is fed to a Multi-domain collaborative Filter (MDCF) based preprocessing process. Then the preprocessed data's are fed to Scaling-Basis Chirplet Transform (SBCT) for extracting features such as tone quality and loudness variations. The extracted features are fed to HGRNN to classify the piano playing performance such as good, normal and poor. The weight parameters of HGRNN are optimized using TDO. The proposed PPVR-HGRNN-TDO is implemented in python, effectiveness assessed by several performance metrics such as accuracy, precision, specificity, sensitivity, F1-score and ROC. The gained results of the proposed PPVR-HGRNN-TDO method attains higher accuracy 20.46%, 23.54%, 35.58%, higher precision 33.56%, 21.72%, 33.97% and higher sensitivity 32.54%, 22.76%, 36.97%. The proposed method shows better results in all existing systems like Dual Channel Convolutional Neural Network (DCCNN),Convolutional Neural Network (CNN), and Back Propagation Neural Network with Genetic Algorithm (BPNN-GA). From the result it is concludes that the proposed PPVR-HGRNN-TDO method based accuracy is higher than the existing methods.

Keywords: Piano Performance, MIDI dataset, Chirplet Transform, Preprocessing, Piano Classroom, Virtual Reality, Real Time Interaction.

I. INTRODUCTION

For college students majoring in preschool education, mastering the piano is essential. A cutting-edge teaching strategy founded on modern educational ideas is piano informatization in the classroom [1, 2]. By utilizing information technology, it expands the scope of the curriculum, eliminates time, place, and location constraints associated with traditional classroom instruction, and achieve independent learning by fully mobilizes students' subjective initiative. One innovative approach to teaching is piano information-based learning in the classroom [3-5]. There aren't many studies that use this method to assess the quality of instruction. Simultaneously, a multitude of interference indicators are available to regulate the quality of instruction in a piano information-based classroom. These indicators are very relevant, mutually exclusive, and comprehensive in terms of evaluation [6–8]. Integrity is a critical factor in evaluating the exceptional influence of piano knowledge classroom instruction evaluation.

Before graduating and becoming kindergarten teachers, they must first acquire a particular degree of piano playing ability through courses [9]. These students are also novices at the piano, thus it is challenging for them to do well in these classes. When learning a tune in class, beginners require continuous practice, as well as immediate and tailored feedback [10, 11]. Furthermore, it takes time and effort to manually score each student's performance on the live exam simulations known as the midterm and final exams [12]. Furthermore, at the start of each lesson, manually grading each student's work is challenging when there is a big course roster [13]. The pianist's real performance during a performance is greatly influenced by the many emotional activities and expressions of the circumstance of the player's performance, as well as psychological challenges like the intensity and horror of the scenario [14].

The information-based piano teaching method, which is centered on in-depth learning, was developed to address the problem of enhancing the efficacy of information-based instruction,innovates the teaching mode and produces the piano information-based classroom [15–17]. It can aid students in understanding the foundational theory and piano performance methods [18]. Modern multimedia piano instruction can also benefit from a complete embrace of new ideas, as traditional piano instruction is no longer adequate to suit the needs of students and the demands of the times. [19].

[1*]Lecturer, Music College of Guangxi Arts University, Nanning, Guangxi, 530022, China
*Corresponding author e-mail: miumiueyes335@163.com

Using a sample classroom design that integrates the most recent educational principles of teaching under network and multimedia technologies, the four piano classroom teaching styles are investigated, analyzed, and contrasted. When used in the classroom, the advantage of teaching increases learning efficiency [20]. The numerous emotional activities and expressions of the pianist's performance, as well as mental difficulties like the intensity and terror of the scene, all have a significant impact on the pianist's actual performance while they are performing thanks to the current virtual reality-based technology. This work attempts to address the shortcomings of current models by providing a novel method for performing on the piano that makes use of the Tasmanian Devil Optimization (TDO)-optimized Hierarchically Gated Recurrent Neural Network (HGRNN).

A. Contribution Statement

- Predicting real time interaction of piano playing performance using Hierarchically Gated Recurrent Neural Network optimized with theTasmanian Devil Optimization (PPVR-HGRNN-TDO).
- The data’s are gathered from the MIDI dataset. Then the data’s are fed to pre-processing.
- Using a Multi-domain collaborative Filter eliminates the noise of input data in the pre-processing segment.
- The pre-processing output is fed into Scaling-Basis Chirplet Transform for extracting texture features such as tone quality and loudness variations.
- The extracted features are fed to Hierarchically Gated Recurrent Neural Network, it classify the piano playing performance such as good, normal and poor.
- The performance of the proposed approach is validated through python platform and compared with other existing techniques.
- After the proposed PPVR-HGRNN-TDO technique is implemented, performance metrics including ROC, F1-score, accuracy, precision, specificity, and sensitivity are examined.

The following is the arrangement of the remaining sectors of this paper: Sector 2 looks at a literature review; sector 3 explains the proposed approach; Sector 4 provides a discussion and results; and sector 5 concludes.

II. LITERATURE SURVEY

Many studies on virtual reality-based piano performance have already been published in literature; some new studies are disclosed below.

Wang et al. [21] have developed For novices, there were two distinct audio-based techniques for evaluating piano performance. Three steps comprised the first sequential and modular system: CNN for auditory feature extraction, DTW for matching, and performance score regression. CNNs and the attention mechanism were both part of the 2nd system, which was an end-to-end architecture. To predict a performance score, it takes as input two sequences of audio features. We used our new open-access YCU-PPE-III, which includes over 2000 piano audio pieces recorded in several actual exam sessions, to test two of the offered strategies.

Feng [22] have introduced Computers are used in VR technology to fully put on the real world. Virtual reality has many uses and has been incorporated into college education extensively. A Virtual reality -based music teaching system was established in order to enhance the usage of computer-aided instruction technology in the field of music education. The earliest hardware platforms used in the development of the virtual piano were an HTC Vive kit and a Leap Motion sensor mounted on a helmet.The software platforms used were Unity3D, relevant Leap Motion and SteamVR add-ons. In particular, the user's gesture instruction input was recorded using the DCCNN. Video gesture commands and image feature information were extracted by the dual-size convolution kernel and then recognized by the DCCNN. Following the retrieval of the temporal and spatial data, optical flow images and RGB color pattern images were supplied into the DCCNN.

Wang [23] have presented The curriculum objectives, education and training objectives, and classroom teaching objectives made up the three categories into which the teaching objectives were divided. It was decided to construct an interactive classroom, learning area, teaching platform, resource platform, and cloud application platform for piano knowledge. To develop a special teaching mode for piano information classrooms, the prior teaching mode was optimized. A model of a hierarchical structure for each assessment index was created using the analytical hierarchy technique,The index system for evaluating the quality of piano instruction in classrooms was constructed using it. For every assessment index, a hierarchical structure model was created using the hierarchical analysis method. The nine-digit scale method was used to create the judgment matrix. The need of evaluating the quality of piano instruction in the classroom was determined after the evaluation matrix's correctness was proven. In the new mode, the deep learning algorithm uses a GA to maximize the weight and threshold of a BP neural network.

Molero et al. [24] have introduced the music students, particularly novices, frequently lack motivation and become discouraged when their musical skills do not enhance despite repeated practice. In such cases, students frequently drop out of music school. In order to overcome this challenge, gamification and mixed reality were used to create a novel approach to inspiring music students. The multimedia application HoloMusic XP, which aids children in learning the piano and music, has validated this method. HoloMusic XP's underlying architecture was created and expanded to accommodate new musical ideas. The use of mixed reality can help new students get over their sometimes steep learning curve, and visual metaphors can help make complex musical concepts easier to understand. Teachers and students assessed the system in real-world settings to determine its usefulness and usability.

Niu [25] have suggested The creation of novel pedagogical strategies and the general extension of educational roles are encouraged by contemporary teaching modes. a penetration analysis of multimedia and complex networks in piano instruction and performance. In order to categorize, analyze, research, and evaluate piano teaching cases, A few typical teaching scenarios that follow modern learning theories, instructional specifications, and piano discipline features were eliminated. The complex network and multimedia technology were utilized for this purpose. Build several network teaching scenarios for piano teachers to make use of, research, and share the benefits while addressing the drawbacks. Enhance the initial instructor-led, one-class teaching strategy by implementing a new teaching model. Multimedia and network technology, in particular, supported the new teaching mode by enabling piano instruction to move toward personalized learning. To find out more about the effects of multimedia piano instruction, questionnaire surveys were administered to teachers and students using this method.

Chen [26] have developed Its use in the sphere of music instruction has grown as well; more effective and sophisticated teaching strategies are now being used in music schools. Virtual reality (VR) technology has been a new tool for scientific research and instruction in education reform. Virtual reality-based teaching systems have been applied to a variety of instructional settings. Consequently, in order to build the VR-based music education platform, a system that used camera setup, bread capture, packing capture, and model embedding was created using virtual reality (VR). The creation of several virtual components can effectively stimulate the singer's senses and better achieve the objectives of audience involvement.

Liu [27] have suggested There were many different chess styles taught in piano lessons, but there were no all-encompassing, methodical, scientific teaching strategies. It fails to meet the current development needs of piano education and draws attention to a number of pedagogical concerns. Nonetheless, the piano scoring system can help pianists somewhat replace their professors' support. The study examined the signal qualities of musical performance, constructed a scoring model for piano performances using BP neural network technology and Big Data, and selected well-known compositions to showcase the efficacy of the system.

III. PROPOSED METHODOLOGY

In this section, the real time interaction for piano playing performance using hierarchically gated recurrent neural network optimized with the Tasmanian devil optimization s discussed. The MIDI dataset is where the input photos are first acquired. The datas are then fed into preprocessing. Preprocessing involves removing noise using multi-domain collaborative Filtering. After that, the preprocessing output is given into the Scaling-Basis Chirplet Transform for extracting texture features such as tone quality and loudness variations. And then, HGRNN is used for classification it is performed for classifying the piano playing performance such as good, normal and poor. Using TDO, the weight parameters of the HGRNN are optimized. Block schematic of the proposed PPVR-HGRNN-TDO method is illustrated in Fig 1. The detail description of proposed methodology are given below,

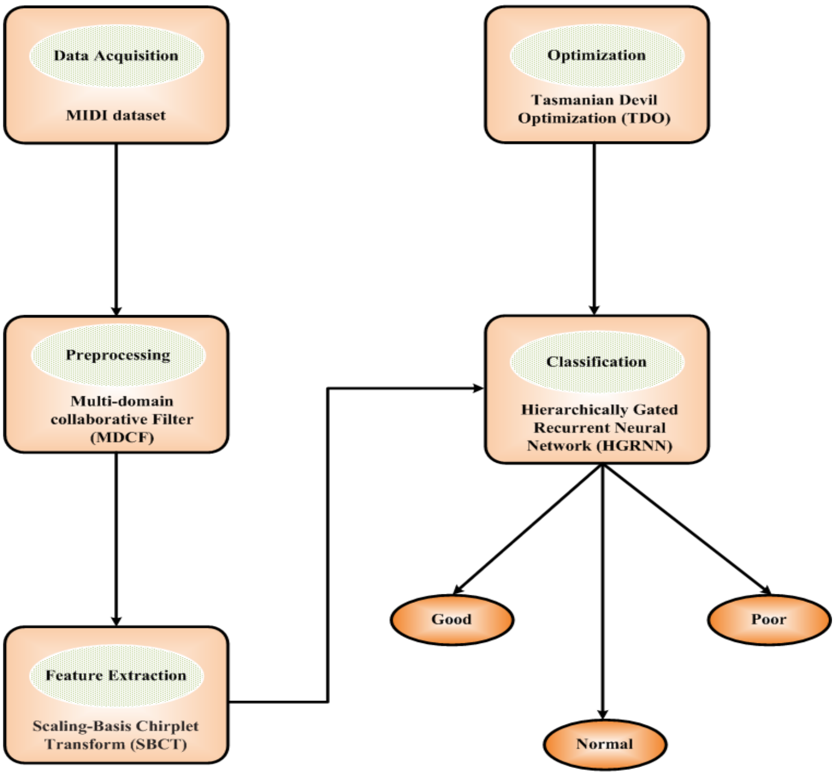

Fig 1: Block schematic of the proposed PPVR-HGRNN-TDO method

A. Data Acquisition

The MIDI dataset, which includes 74 performance recordings totaling 3 hours and 8 minutes and 16 separate tracks (eight piano duets) performed by two pianists, one male and one female, is where the piano playing-based data are collected. Eight piano duets were assigned to the two players, the primo and secondo parts going to each [28]. After that, each player performed the eight recordings one to seven times in order to create a variety of expressive styles, such as normal, exaggerated, etc. To guarantee that the players have adequate visual expressiveness for interactions, the primo and secondo are recorded jointly at each instance. The MIDI keyboard automatically encodes the key pressing data (pitch, timing, and velocity) into the MIDI standard. Semi-automatically connecting the MIDI stream with the relevant MIDI score data annotated the quantization beat number and the downbeat places for each recording.

B. Pre-processing using Multi-domain collaborative Filter (MDCF)

In this phase, Multi-domain collaborative Filter (MDCF) pre-processes the input data, which is used to reduce noise from the input data and enhance the cleanliness of the data.

The structure of a complex signal $z(t)$, with N factors in M domains, is the same [29]. Assume that p_i^j represents the jth factor in the multi-domain's ith domain $J \in \{1,2,..,n_j\}$, n_i is the entire quantity of factors in the ith domain, and d_i signifies every distinct domain with $i \in J := \{1,2,..,M\}$. It appears that $\sum_{i \in I} n_i = N$.

Next, in order to represent multi-domain spaces, we define a multi-domain character vector as

$$V = \left[p_1^1 \dots p_1^{n1} \dots p_M^1 \dots p_M^{n_M}\right] \tag{1}$$

The multi-domain expressions for each known parameter are distinct, allowing V to provide a complete description of both the interference and the target signal.

The target signal plus several multi-domain interferences make up the received signal model $z(t)$, which also be expressed as follows:

$$z(t) = a_{0,\xi}\xi(t) + \sum_{l=1}^{L} a_{0,l}\mu_l(t) + n(t) \tag{2}$$

Where $\mu_l(t)$ denotes low interference and $\xi(t)$ denotes the target signal. For the *l*thinterferingand the target signal, correspondingly, $a_{0,\xi}$ and $a_{0,l}$ are denoted as multi-domain collaborative steering vectors.

For several domains, co-oblique projection operator is described as

$$Ea_{0,\xi}A_{o,\mu} = a_{0,\xi}\left(a_{0,\xi}^{H}P_{Ao,\mu}^{\perp}a_{0,\xi}\right)^{-1}a_{0,\xi}^{H}P_{Ao,\mu}^{\perp} \tag{3}$$

Where, $A_{0,\mu} = \left[a_{0,1}, a_{0,2} ... a_{0,L}\right]$ is made up of each interference's multi-domain collaborative steering vectors. $a_{0,l} = a_{1,l} \otimes a_{2,l} \otimes ... \otimes a_{m,l}$ is the*l*th interference's multi-domain collaborative steering vectors.

The MDC filter, is defined as follows to obtain the amplitude of target signal:

$$w_0 = \left(F_0 Ea_{0,\xi}A_{0,\mu}\right)^H \tag{4}$$

Where, $F_0 = \left(a_{0,\xi}^{H}a_{0,\xi}\right)^{-1}a_{0,\xi}^{H}$. Then the filter output is provided as follows:

$$\xi'(t) = w_0^H z(t) = w_0^H a_{0,\xi}\xi(t) + w_0^H\left(\sum_{l=1}^{L} a_{0,l}\mu_l(t) + n(t)\right) \tag{5}$$

The aforementioned equation demonstrates that for high SNRs, interference is totally suppressed, whereas the multi-domain recovery of the target signal is possible without distortion.

The target signal's steering vectors and those of any interference can be shown as $a_{i,\xi}$ $(i = 1,2,..,m)$, $A_{i,\mu} = \left[a_{i,1}, a_{i,2} ..... a_{i,L}\right]$the multi-domain's ith individual domain, the appropriate individual domains' weight steering vectors for filtering are provided as

$$w_i = \left(\left(a_{i,\xi}^{H}a_{i,\xi}\right)^{-1}a_{i,\xi}^{H}Ea_{i,\xi}A_{i,\mu}\right)^H \tag{6}$$

The operator for oblique projection by $a_{i,\xi}$ and $A_{i,\mu}$ is denoted by $Ea_{i,\xi}A_{i,\mu}$. Next, each known w_i can be used to specify a multi-domain's weight steering vector, which can thereafter be further rewritten as

$$w_0^H = w_m^H \cdot w_{m-1}^H \cdots w_1^H \tag{7}$$

When only the space and polarization domains are taken into account, the vector for weight steering w_{sp} of the collaborative space polarization domain will be regarded as w_0^H. Each specific domain's weight steering vectors (i.e., w_s and w_p) and w_{sp}'s mathematical relationship,

$$w_{sp}^H = w_m^H \cdot w_{m-1}^H \cdots w_1^H \tag{8}$$

Thus this method removes the noise from the input data.Lastly, the pre-processed output is given towards feature extraction.

C. Feature Extraction using Scaling-Basis Chirplet Transform (SBCT)

Here, SBCT is tasked with extracting features from the pre-processed data.

The TF analysis method, commonly referred to as the SBCT, was created by expanding the traditional chirplet transform [30]. This method allows the chirp rate to vary with frequency and time by employing a replacement kernel function to scale the TF basis at and around the appropriate time center. For a given signal $x(u) \in L^2(R)$, the CT can be expressed as,

$$CT(f, t_c) = \int_{-\infty}^{+\infty} s(u)h(u - t_c)\exp(-j2\pi\varphi(f, u, t_c))du \tag{9}$$

Here, the analytic signal of $x(u)$produced by the Hilbert transform is represented by the $s(u)$. H denotes a real window function that is non-negative, symmetric, and normalized. It is typically represented by a Gaussian function with the following formula:

$$h_\sigma(u) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2\right) \tag{10}$$

Where standard deviation is shown by σ.

A new kernel phase function with chirp rate as its second derivative, which is dependent on the integration variable, frequency center, and time center:

$$C = d\varphi'/du = c(f, t_c, u) = -\tan(\theta) \tag{11}$$

Here u is denoted as the variable of integration and f and t_c stand for the time and frequency centers, correspondingly. Among these, u guarantees that, across the whole window length, the IF trajectories and the TF basis align, θ is the TF basis's rotating angle, and f and t_c make sure that, for various time centers and frequency components, the IF trajectories and the TF base match. correspondingly. Taking into account the connection between the chirp rate and the TF basis's rotating angle, θ changes as a function of f, t_c, and u.

To achieve the goal, the initial basis is scaled in order to create a new phase function φ_s that satisfies the previously indicated requirements:

$$\varphi_s(f, u, t_c, a_1, a_2, \ldots, a_n) = f \times \left(u + \sum_{k=1}^{n} a_k (u - t_c)^{1+k} \right) \tag{12}$$

Here the parameters that need to be found are $a_1, a_2, \ldots, a_n$. The first and second derivatives that correspond to them can be expressed as,

$$\varphi_s' = \frac{d\varphi_s}{du} = f \times \left(1 + \sum_{k=1}^{n} (1+k) a_k (u - t_c)^k \right) \tag{13}$$

$$-\tan(\theta) = \frac{d\varphi_s'}{du} = f \times \left(\sum_{k=1}^{n} (1+k) k a_k (u - t_c)^{k-1} \right) \tag{14}$$

The IF of the phase function φ_s is represented by φ_s'.

Finally, the SBCT can be stated as

$$SBCT(f, t_c, a_1, a_2, .., a_n) = \int_{-\infty}^{+\infty} s(u) h(u - t_c) \exp\left(-j2\pi f \times \left(u + \sum_{k=1}^{n} a_k (u - t_c)^{1+k} \right) \right) du \tag{15}$$

Here, $s(u)$ is denoted as the analytical signal of $x(u)$, $a_1, a_2, .., a_n$ are parameters, u is the integration variable.

The extracted features are tone quality and loudness variations. Then the extracted output is given to classification phase.

D. Classification using Hierarchically Gated Recurrent Neural Network (HGRNN)

Classification of real time interaction of piano playing utilizing Hierarchically Gated Recurrent Neural Network (HGRNN) was discussed in this section [31]. A hierarchically gated recurrent neural network features multiple stacked layers, which each has a channel mixing module (GLU) and a token mixing module (HGRU). A basic gated linear recurrent linear unit is defined as follows:

HGRU exploration: A basic linear recurrent layer that is gated is described as follows:

$$F_T = Sigmoid(Z_T M_F + A_F) \in D^{1\times F} \tag{16}$$

$$J_T = Sigmoid(Z_T M_J + A_J) \in D^{1\times F} \tag{17}$$

Here F_T and J_T represent forget and input gates respectively.

Complex valued recurrence: To accomplish element-wise linear recurrence in linear RNNs with static decay rates, eigen decompositions of the recurrent weight matrix are commonly used. The expressiveness of the model is diminished when real-valued eigenvalues are the only ones used, as this restricts the recurrent weight matrix's symmetric range. We parameterize the real and imaginary sections of the input G_T independently in the manner described below:

$$\mathrm{Re}(G_T) = SiLU(Z_T M_{GR} + A_{CR}) \in D^{1\times F} \tag{18}$$

$$\mathrm{Im}(G_T) = SiLU\left(Z_T\, M_{Gj} + A_{Cj}\right) \in D^{1\times F} \tag{19}$$

Lower bound on forget gate values: Only the magnitude argument λ_T affects how much information is remembered. In practical terms, parameterize lower bounds separately or for every hidden state, here E is the count of layers. The following computations are written as follows, assuming that the layer index is P :

$$Q = \left(Soft\max\left(\Gamma, \dim = 0\right) \in D^{E\times F}\right) \tag{20}$$

$$\gamma^P = \left[Cumsum\left(Q, \dim = 0\right)\right] p \in D^{1\times F} \tag{21}$$

Lastly, the P-th layer's λ_T is parameterized as follows:

$$\mu_T = Sigmoid\left(Z_T\, M_\mu + A_\mu\right) \in D^{1\times F} \tag{22}$$

To achieve the same forget rate value $\bar{\gamma}$ closed to one, μ_T will be pushed away from the sigmoid activation function's saturated regions.

$$\mu_T = \frac{\bar{\gamma} - \gamma^p}{1 - \gamma^p} < \bar{\gamma} \tag{23}$$

Trying inputting and forgetting gates: Leaky units can be useful for reducing the amount of parameters. These units have a close relationship to exponential moving averages and the discretization of continuous time systems, and they have shown empirical usefulness.

$$E_T = \lambda_T\, \Theta \exp\left(j\theta\right) \Theta\, E_{T-1} + \left(1 - \lambda_T\right) \Theta\, G_T \in \subset^{1\times F} \tag{24}$$

Here Θ denotes the element wise product.

Output gates and projection: It has been demonstrated that adding gates to the recurrence layer's output enhances state space models. Before completing the output projection to obtain HGRU, an output gate is used in the following manner.

$$k_T = Sigmoid\left(M_k\, Z_T + A_k\right) \in D^{1\times 2r} \tag{25}$$

The optimal signal is predicted by proposed HGRNN technique. Finally, the HGRNN classifies the good, normal and poor of the piano playing performance.

E. Optimization using Tasmanian Devil Optimization (TDO)

The TDO is used to determine the ideal parameters for the HGRNN classifier. A carnivorous wild mammal belonging to the Dasyuridae family, the Tasmanian devil is found on the island state of Tasmania [32]. Although Tasmanian devils may hunt, they are opportunistic eaters who prefer to eat carrion when it is available. The Tasmanian devil sustains itself in two ways. In the first tactic, a Tasmanian devil consumes carrion as it comes across it. The second approach is to attack its target in order to hunt and feast on it. Indeed, the Tasmanian devil's hunt for food in numerous regions resembles the exploration index, which is utilized in the process of optimizing search space to locate the ideal location. The pursuit of the Tasmanian devil of its prey inside a narrow area, on the other hand, is similar to the local search's exploitation index, which similarly seeks to find the best response. This implies that mathematical simulations of Tasmanian devil foraging approaches will likely result in the development of an optimizer capable of coming up with the best answers to optimization issues. The flow chart of TDO algorithm is illustrated in Fig 2.

Step 1: Initialization

Set the inputs' initial values. here the input parameter are weight parameter of HGRNN. Which is denoted as k_T .

Step 2: Random Generation

After setup, the random vectors generate the input parameters at random.

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,d} \\ \cdots & \cdots & \cdots & \cdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,d} \end{bmatrix} \tag{26}$$

Where, n is denoted as the initialized parameter, and d is denoted as the dimension of the desirable variables.

Step 3: Fitness Calculation

The fitness is selected based on the objective function.

$$F = Optimize\ (k_T) \tag{27}$$

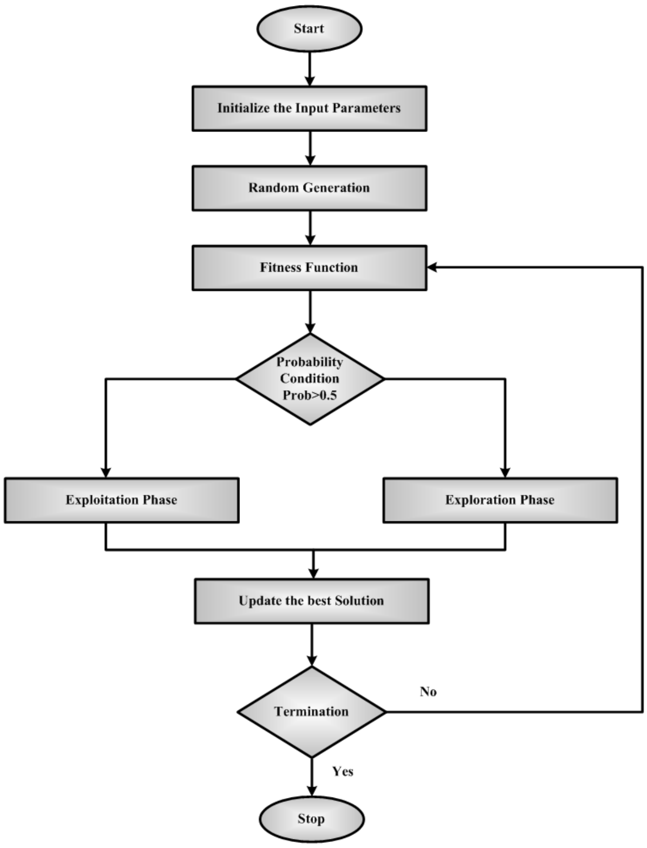

Fig 2: Flow chart of TDO algorithm

Step 4: Exploration Phase

Where other population fellows are located in the search space represents each Tasmanian devil's carrion site in the TDO concept. The following equation represents a random selection of one of these scenarios, wherein the ith Tasmanian devil chooses to target the kth person in the population as his carrion. I is the opposite, and K must be selected at random from 1 to N.

$$C_i = X_k, \qquad i = 1,2,.....,N,\ k \in \{1,2,....,N \mid k \neq i\} \tag{28}$$

Here, C_i is denoted as the chosen carrion by the ith Tasmanian devil. The picked carrion determines the Tasmanian devil's new location within the search space. In this strategy's motion simulation, the Tasmanian devil advances toward carrion if its goal function value is higher; else, it moves away from it. The following equation models the Tasmanian devil's moving strategy.

$$x_{i,j}^{new,S1} = \begin{cases} x_{i,j} + r.(c_{i,j} - I.x_{i,j}), & Fc_i < F_i; \\ x_{i,j} + r.(x_{i,j} - c_{i,j}), & otherwise \end{cases} \tag{29}$$

The values for the jth variable and the first strategy's result are indicated by $x_{i,j}^{new,S1}$, the new status of the Tasmanian devil; Fc_i, the objective function value of the chosen carrion; r, the arbitrary value in the range [0,1], and I, the arbitrary value, which may be either 1 or 2.

Step 5: Exploitation Phase

Tasmanian fallen angels' second mode of survival is to chase and consume prey. When Tasmanian devils attack, they exhibit two distinct behaviors. In the first phase, it searches its surroundings, selects a target, and attacks. The kth person in the population is randomly chosen to be the prey using a natural arbitrary value between 1 and the reverse of i. To replicate the process of choosing prey, use the following equation.

$$P_i = X_k, \qquad i = 1,2,.....,N,\ k \in \{1,2,....,N \mid k \neq i\} \tag{30}$$

Where, P_i is denoted as the prey that the Tasmanian devil has selected.

The Tasmanian devil's new position is calculated after the prey's location is established. The Tasmanian devil moves toward its chosen prey in order to calculate its new location, supposing that the target capability is worth more. In any case, it causes some displacement from the initial position. The equation below illustrates how this process is modeled.

$$x_{i,j}^{new,S2} = \begin{cases} x_{i,j} + r.\left(p_{i,j} - I.x_{i,j}\right), & Fp_i < F_i; \\ x_{i,j} + r.\left(x_{i,j} - p_{i,j}\right), & otherwise \end{cases} \tag{31}$$

Where, $x_{i,j}^{new,S2}$ is denoted as the second strategy's basis for the Tasmanian's new status Fp_i is denoted as the chosen prey's objective unction value.

Step 6: Chasing and Updating Strategy

Hunting down prey near the attack site is analogous to conducting a local search of the search space. The Tasmanian devil's actions really show how the TDO could be applied to more effectively find potential solutions. The Tasmanian devil imitates this chase by following its prey around the attack site. The Tasmanian devil position is currently regarded to be the neighborhood's hunting hub. The radius of the neighborhood can be determined using the method below to determine how far the Tasmanian devil travels in pursuit of its meal.

$$R = 0.01\left(1 - \frac{t}{T}\right) \tag{32}$$

Where, *t*signifiesthe counter for iterations, *R* is denoted as the attack point's surrounding radius, and *T* is denoted as the maximum iteration number.

The Tasmanian devil's new location, as stated in the equation below, can be established using the pursuit process in this area.

$$x_{i,j}^{new} = x_{i,j} + (2r - 1).R.x_{i,j} \tag{33}$$

Where, $x_{i,j}^{new}$ is denoted as the new status of the ith and jth variable of Tasmanian devil.

Step 7: Termination

Check the termination criteria, if it met the condition means optimal solution is obtained, otherwise repeat the process.

IV. RESULT & DISCUSSION

The result of proposed piano playing performance of virtual reality using Hierarchically Gated Recurrent Neural Network optimized with theTDO (PPVR-HGRNN-TDO) is discussed. The Python platform is used to verify the proposed approach's performance, which is then contrasted with other existing techniques. The obtained results of the proposed with PPVR-HGRNN-TDO technique are evaluated with existing techniques like PPVR-CNN, PPVR-DCCNN and PPVR-BPNN-GA methods.

A. Performance Measures

To examine the performance, the performance metrics as accuracy, precision, ROC, F1-score, sensitivity and specificity are determined.

1) Accuracy

It is defined as the ratio of a precise forecast to the entire count of proceedings in the dataset. The following formula (34) is used to calculate accuracy.

$$Accuracy = \frac{(T_P + T_N)}{(T_P + F_P + T_N + F_N)} \tag{34}$$

2) F1-Score

It evaluates the precision of the model on the dataset. It is determined by equation (35)

$$F1Score = \frac{T_P}{\left(T_P + \frac{1}{2}\left[F_P + F_N\right]\right)} \tag{35}$$

3) Precision

Precision is the positive predict value. Precision is compute by following equation (36)

$$\mathrm{Pr}ecison = \frac{T_P}{T_P + F_P} \tag{36}$$

4) ROC

ROC provides an overall performance indicator for the whole probable Classification. ROC is given in eqn (37)

$$ROC = 0.5 \times \left(\frac{T_p}{T_P + F_N} + \frac{T_N}{T_N + F_P} \right) \tag{37}$$

5) Sensitivity

Equation (38) can be employed to define sensitivity.

$$Sensitivity = \frac{T_P}{T_P + F_N} \tag{38}$$

6) Specificity

A metric called specificity is used to assess the percentage of real negatives that are accurately detected. Specificity is calculated using the Following equation (39)

$$Specificity = \frac{T_N}{T_N + F_P} \tag{39}$$

B. Performance Analysis

The simulation results of proposedPPVR-HGRNN-TDO method shows in Fig 3-8. Then the proposed PPVR-HGRNN-TDO method is likened with existing systems likePPVR-CNN, PPVR-DCCNN and PPVR-BPNN-GA respectively. In order to show the effectiveness of the proposed HGRNN classifier with optimization TDO algorithm, evaluation experiment was carried and the results.

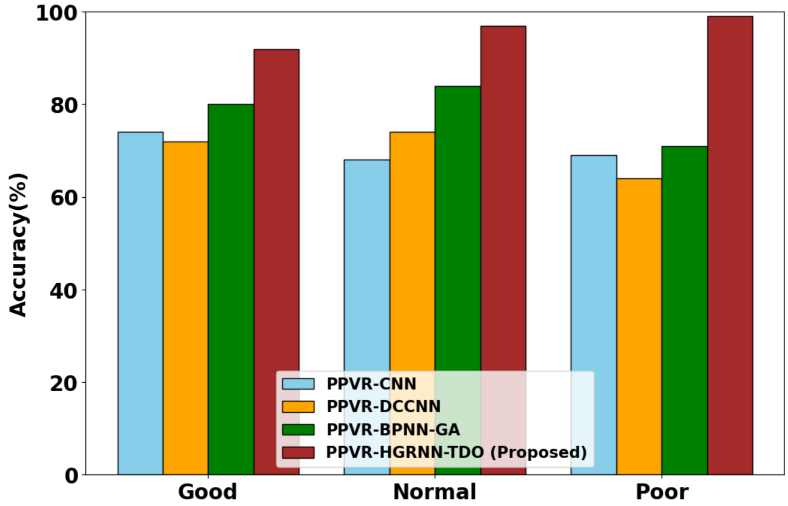

Fig 3: Performance analysis of accuracy

The performance analysis of accuracy is displayed in Fig 3. The performance of the proposed PPVR-HGRNN-TDO technique results in accuracy that are 25.76%, 32.97%, 52.56% higher for the classification of good, 20.46%, 23.54%, 35.58% higher for the classification of normal and 21.45%, 30.76%, 18.43% higher for the classification of poor. When compared to the existing models of PPVR-CNN, PPVR-DCCNN, and PPVR-BPNN-GA, in that order.

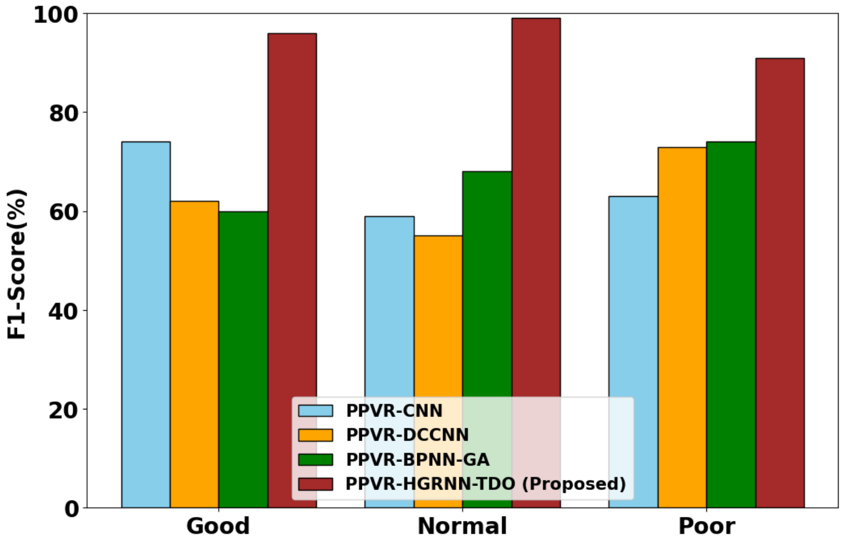

Fig 4: performance analysis of F1-score with proposed and existing methods

The performance analysis of F1-score with proposed and existing methods is revealed in Fig: 4.The performance of the proposed PPVR-HGRNN-TDO technique results in F1-score that are 23.56%, 24.76%, 35.97% higher for the classification of good, 22.46%, 30.58%, 24.54% higher for the classification of normal and 21.45%, 32.76%, 18.43% higher for the classification of poor when evaluated to the existing PPVR-CNN, PPVR-DCCNN and PPVR-BPNN-GA models correspondingly.

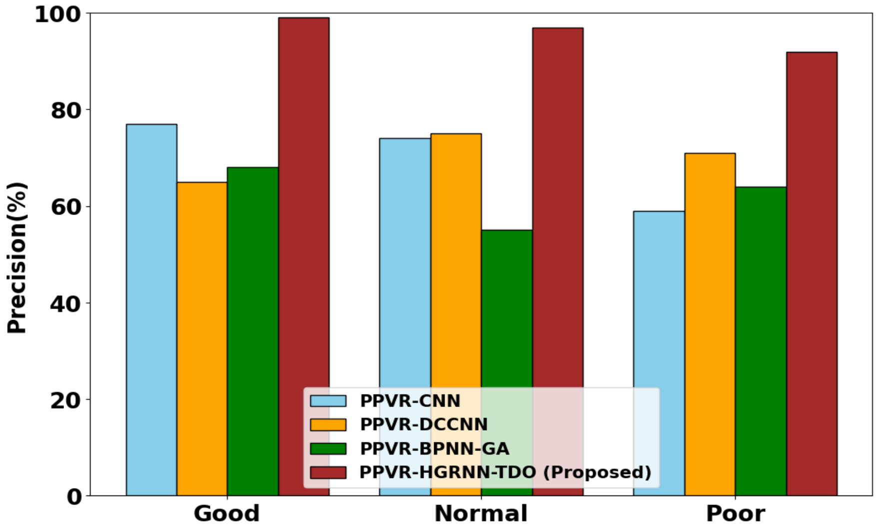

Fig 5: Comparison of precision with proposed and existing method

The comparison of precision with proposed and existing method is illustrated in Fig 5. The performance of the proposed PPVR-HGRNN-TDO technique results in precision that are 33.56%, 21.72%, 33.97% higher for the classification of good, 22.46%, 31.58%, 25.54% higher for the classification of normal and 24.45%, 33.76%, 19.43% higher for the classification of poor when evaluated to the existing PPVR-CNN, PPVR-DCCNN and PPVR-BPNN-GA models respectively.

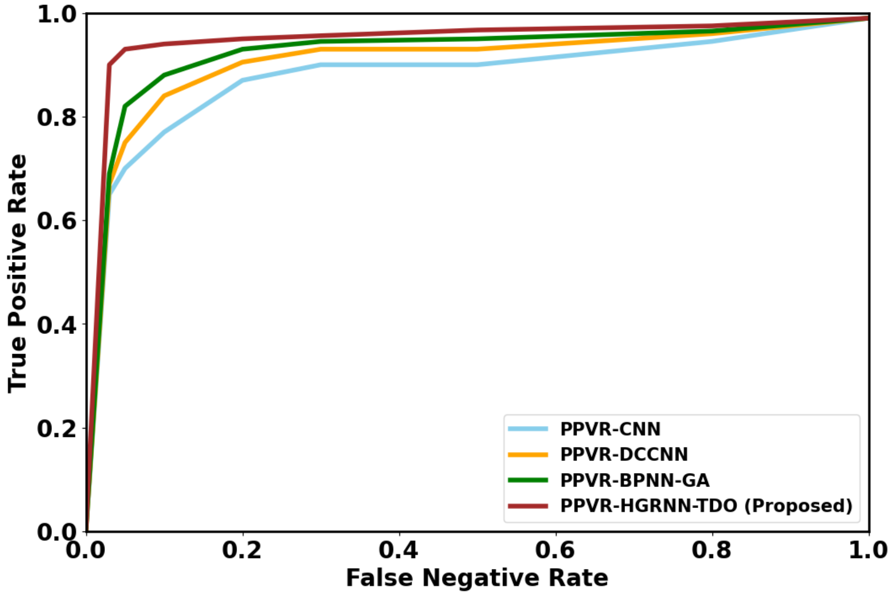

Fig 6: Performance analysis of ROC

The performance analysis of ROC is shown in Fig 6.Each point on the curve represents a different threshold value, and the curve is created by connecting these points. The prediction system performs better the closer the curve is to the top-left corner of the graph.. The higher ROC indicates better performance in distinguishing between positive and negative instances. The proposed PPVR-HGRNN-TDO method the ROC provides high piano playing performance compared with other existing approaches. The existing approaches like PPVR-CNN, PPVR-DCCNN and PPVR-BPNN-GA the ROC become lower compared with the proposed PPVR-HGRNN-TDO technique.

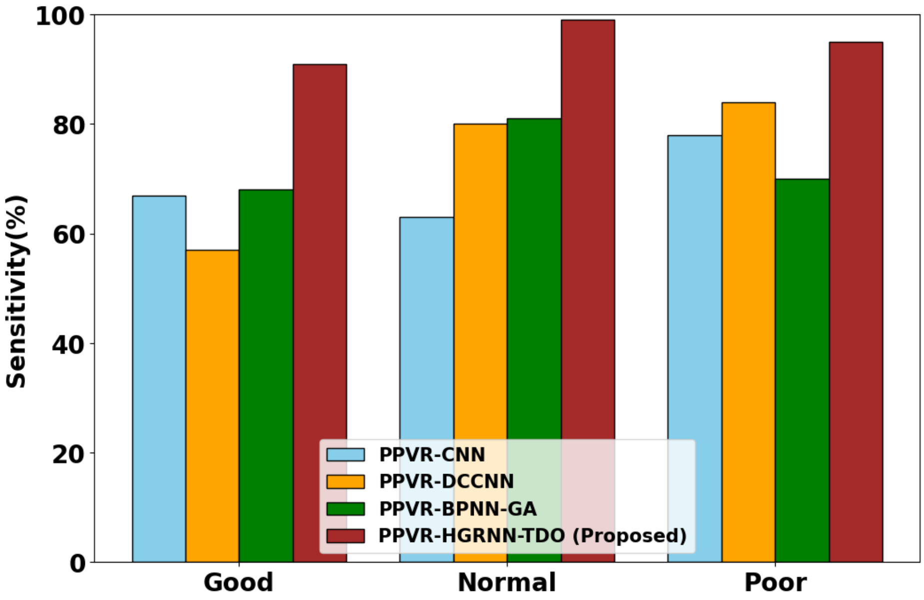

Fig 7: Comparison of sensitivity with proposed and existing method

The comparison of sensitivity with proposed and existing method is illustrated in Fig 7.The performance of the proposed PPVR-HGRNN-TDO technique results in sensitivity that are 32.54%, 22.76%, 36.97% higher for the classification of good, 21.46%, 31.58%, 25.54% higher for the classification of normal and 22.48%, 28.72%, 18.47% higher for the classification of poor, when evaluated to the existing PPVR-CNN, PPVR-DCCNN and PPVR-BPNN-GA models correspondingly.

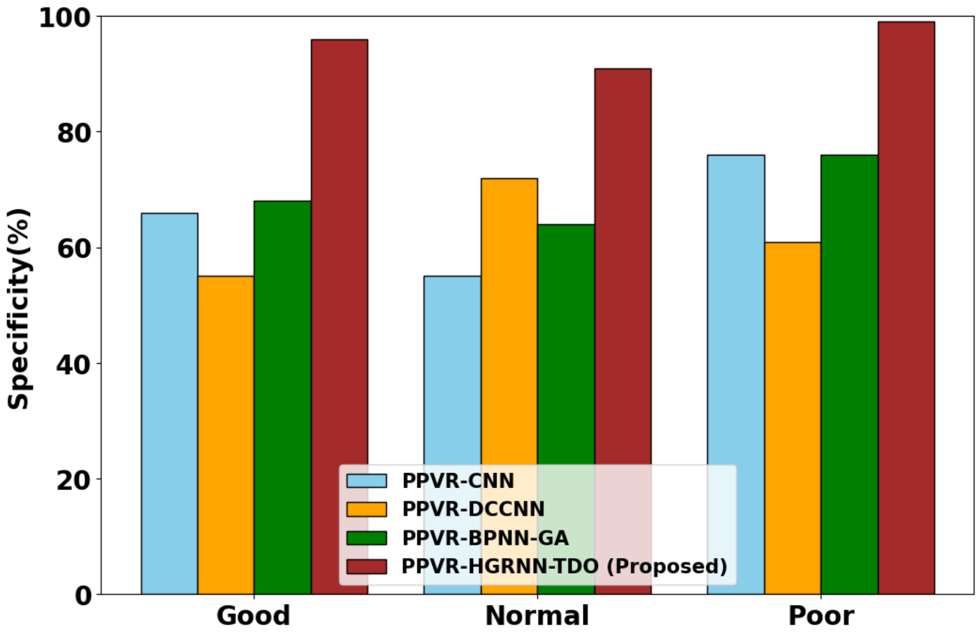

Fig 8: Comparison of specificity with proposed and existing method

The comparison of specificity with proposed and existing method is illustrated in Fig: 8.The performance of the proposed PPVR-HGRNN-TDO technique results in specificity that are 30.72%, 24.89%, 35.92% higher for the classification of good, 21.55%, 35.52%, 29.52% higher for the classification of normal and 20.45%, 27.80%, 20.54% higher for the classification of poor, when evaluated to the existing PPVR-CNN, PPVR-DCCNN and PPVR-BPNN-GA models respectively.

V. CONCLUSION

In this section, piano playing performance of virtual reality using PPVR-HGRNN-TDO method was successfully implemented for classifying good, normal and poor of the piano playing performance. The proposed PPVR-HGRNN-TDO method is executed in the python platform utilizing the dataset of MIDI dataset. The performance of the PPVR-HGRNN-TDO method contains precision, accuracy, sensitivity, F1-score, specificity, and ROC. The proposed PPVR-HGRNN-TDO method attains 25.76%, 32.97%, 52.56% higher accuracy, 33.56%, 21.72%, 33.97% higher precision, 32.54%, 22.76%, 36.97% higher sensitivity, 23.56%, 24.76%, 35.97% higher FI-score and 30.72%, 24.89%, 35.92% higher specificity. The proposed method shows better results in all existing systems like PPVR-CNN, PPVR-DCCNN and PPVR-BPNN-GA. From the outcome it is concludes that the proposed PPVR-HGRNN-TDO method based sensitivity is higher than the existing methods.

References

[1] Bobbe, T., Oppici, L., Lüneburg, L. M., Münzberg, O., Li, S. C., Narciss, S., ...&Muschter, E. (2021). What early user involvement could look like—developing technology applications for piano teaching and learning. *Multimodal Technologies and Interaction*, *5*(7), 38.

[2] Putranda, I. G., Yumna, A. M., Rosmansyah, Y., &Sukmana, Y. (2023). Exploring Audio Processing in Mixed Reality to Boost Motivation in Piano Learning. *IEEE Access*.

[3] Zhang, Y. (2021). Vr technology to the adjustment of piano playing mentality. *Mathematical Problems in Engineering*, *2021*, 1-12.

[4] Cui, K. (2023). Artificial intelligence and creativity: piano teaching with augmented reality applications. *Interactive Learning Environments*, *31*(10), 7017-7028.

[5] Jamal, I. M. (2022). *Piano teaching mobile application with augmented reality* (Master's thesis, OndokuzMayısÜniversitesi, LisansüstüEğitimEnstitüsü).

[6] Yang, Z. Y. (2020). Modern piano teaching technologies: Accessibility, effectiveness, the need for pedagogues. *Ilkogretim Online*, *19*(3).

[7] Pace, K. B. (2022). Eliciting Music Performance Anxiety of Vocal and Piano Students Through the Use of Virtual Reality.

[8] Aufegger, L., &Wasley, D. (2020). Virtual reality feedback influences musicians' physical responses and mental attitude towards performing. *Music and Medicine*, *12*(3), 157-166.

[9] Tian, K. (2022). Construction of Virtual Piano Performance System Based on Visual Gesture Recognition. *Mobile Information Systems*, *2022*.

[10] Kim, S., Park, J. M., Rhyu, S., Nam, J., & Lee, K. (2021). Quantitative analysis of piano performance proficiency focusing on difference between hands. *PloS one*, *16*(5), e0250299.

[11] Kim, H., Ramoneda, P., Miron, M., & Serra, X. (2022). An overview of automatic piano performance assessment within the music education context.

[12] Liu, M., & Huang, J. (2021). Piano playing teaching system based on artificial intelligence–design and research. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 3525-3533.

[13] Rhyu, S., Kim, S., & Lee, K. (2022). Sketching the expression: flexible rendering of expressive piano performance with self-supervised learning. *arXiv preprint arXiv:2208.14867*.

[14] Rhyu, S., Kim, S., & Lee, K. (2022). Sketching the expression: flexible rendering of expressive piano performance with self-supervised learning. *arXiv preprint arXiv:2208.14867*.

[15] Jeong, D., Kwon, T., Kim, Y., & Nam, J. (2019, May). Score and performance features for rendering expressive music performances. In *Music encoding conference* (pp. 1-6). Vienna, Austria: Music Encoding Initiative.

[16] Wang, W., Pan, J., Yi, H., Song, Z., & Li, M. (2021). Audio-based piano performance evaluation for beginners with convolutional neural network and attention mechanism. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 1119-1133.

[17] Yan, L. (2019). Design of piano teaching system based on internet of things technology. *Journal of Intelligent & Fuzzy Systems*, *37*(5), 5905-5913.

[18] Antoniadis, P., Paschalidou, S., Duval, A., Jégo, J. F., &Bevilacqua, F. (2022, May). Rendering embodied experience into multimodal data: concepts, tools and applications for Xenakis' piano performance. In *Xenakis 22: Centenary International Symposium*.

[19] Micheloni, E., Tramarin, M., Rodà, A., &Chiaravalli, F. (2019). Playing to play: a piano-based user interface for music education video-games. *Multimedia Tools and Applications*, *78*, 13713-13730.

[20] Shuo, C., & Xiao, C. (2019). The construction of internet+ piano intelligent network teaching system model. *Journal of Intelligent & Fuzzy Systems*, *37*(5), 5819-5827.

[21] Wang, W., Pan, J., Yi, H., Song, Z., & Li, M. (2021). Audio-based piano performance evaluation for beginners with convolutional neural network and attention mechanism. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 1119-1133.

[22] Feng, Y. (2023). Design and research of music teaching system based on virtual reality system in the context of education informatization. *Plos one*, *18*(10), e0285331.

[23] Wang, X. (2022). Piano information teaching mode based on deep learning algorithm. *Wireless Communications and Mobile Computing*, *2022*.

[24] Molero, D., Schez-Sobrino, S., Vallejo, D., Glez-Morcillo, C., &Albusac, J. (2021). A novel approach to learning music and piano based on mixed reality and gamification. *Multimedia Tools and Applications*, *80*(1), 165-186.

[25] Niu, Y. (2021). Penetration of multimedia technology in piano teaching and performance based on complex network. *Mathematical Problems in Engineering*, *2021*, 1-12.

[26] Chen, W. (2022). Research on the design of intelligent music teaching system based on virtual reality technology. *Computational Intelligence and Neuroscience*, *2022*.

[27] Liu, X. (2022). Research on piano performance optimization based on big data and BP neural network technology. *Computational Intelligence and Neuroscience*, *2022*.

[28] Kong, Q., Li, B., Chen, J., & Wang, Y. (2020). Giantmidi-piano: A large-scale midi dataset for classical piano music. *arXiv preprint arXiv:2010.07061*.

[29] Mao, X. P., Yang, Y. L., Hong, H., & Deng, W. B. (2016). Multi-domain collaborative filter for interference suppressing. *IET Signal Processing*, *10*(9), 1157-1168.

[30] Li, M., Wang, T., Chu, F., Han, Q., Qin, Z., &Zuo, M. J. (2020). Scaling-basis chirplet transform. *IEEE Transactions on Industrial Electronics*, *68*(9), 8777-8788.

[31] Qin, Z., Yang, S., &Zhong, Y. (2024). Hierarchically gated recurrent neural network for sequence modeling. *Advances in Neural Information Processing Systems*, *36*.

[32] Dehghani, M., Hubálovský, Š.,&Trojovský, P. (2022). Tasmanian devil optimization: a new bio-inspired optimization algorithm for solving optimization algorithm. *IEEE Access*, *10*, 19599-19620.