Assessment of Surgical Expertise in Virtual Reality Simulation by Hybrid Deep Neural Network Algorithms

Bekir Karlik bkarlik@hotmail.com
Neurosurgical Simulation and Artificial Intelligence Learning Centre
Department of Neurology and Neurosurgery
McGill University, Montreal, QC, Canada

Recai Yilmaz recai.yilmaz@mail.com
Neurosurgical Simulation and Artificial Intelligence Learning Centre
Department of Neurology and Neurosurgery
McGill University, Montreal, QC, Canada

Alexander Winkler-Schwartz awinklerschwartz@gmail.com
Neurosurgical Simulation and Artificial Intelligence Learning Centre
Department of Neurology and Neurosurgery
McGill University, Montreal, QC, Canada

Nykan Mirchi nykan.mirchi@mail.mcgill.ca
Neurosurgical Simulation and Artificial Intelligence Learning Centre
Department of Neurology and Neurosurgery
McGill University, Montreal, QC, Canada

Vincent Bissonnette vincent.bissonnette@mail.mcgill.ca
Division of Orthopaedic Surgery
Montreal General Hospital
McGill University, 1650 Cedar Avenue, Montreal
Quebec, Canada. H3G 1A4.

Nicole Ledwos nicole.ledwos@mail.mcgill.ca
Neurosurgical Simulation and Artificial Intelligence Learning Centre
Department of Neurology and Neurosurgery
McGill University, Montreal, QC, Canada

Rolando Del Maestro rolando.delmaestro@mcgill.ca
Neurosurgical Simulation and Artificial Intelligence Learning Centre
Department of Neurology and Neurosurgery
McGill University, Montreal, QC, Canada

Abstract

The utilization of simulation in surgical resident assessment and training allows quantitation of technical skills in risk-free environments. This transformation of current training paradigms which involve the development and application of high-fidelity virtual reality simulators may play an important role in future surgical educational curricula. With the implementation of artificial intelligence (AI) virtual reality simulators have the potential to advance the understanding, training and assessment of psychomotor performance of surgical expertise. In this study, we apply four hybrid deep neural network algorithms called Fuzzy Clustering Neural Networks (FCNN-1, FCNN-2, FCNN-3, and FCNN-4). The proposed study consists of four hybrid deep neural networks methods called Fuzzy Clustering Neural Network (FCNN) were employed to differentiate surgical expertise. Raw data was obtained from virtual reality tumor resection studies utilizing the NeuroVR simulator platform. The performance of neurosurgeons, senior residents, junior residents and medical students was assessed using a series of metrics. For two proposed algorithms (FCNN-3 and FCNN-4), we achieved a further improvement in classification accuracy of 100% correctly classifying the level of expertise of all participants. The other 2 models were also better accuracies than previous studies. The results demonstrate that the four hybrid FCNN algorithms utilized have increased the accuracy of classification compared to previous studies utilizing the same dataset.

1. INTRODUCTION

Virtual reality technology is being increasingly implemented in many areas of medicine and surgery. The advantage of utilizing this technology is to provide safe simulated environments for healthcare trainees and professionals to develop their skillset and knowledge. In addition, surgical education can be enhanced by the use of virtual reality surgical simulators by allowing an objective assessment of performance and improved evaluation and training of skills. These studies have focused on defining novel metrics, classification of expertise and describing objective proficiency benchmarks (Fried et al., 2019; Gelinas-Phaneuf, and Del Maestro, 2013; Alotaibi et al., 2015).

A series of machine learning algorithms have been employed to assess surgical skills performance on a variety of virtual reality simulators. Studies (Moorthy et al., 2003; Megali et al., 2006; Liang et al., 2011; Rhienmora et al., 2011) have used Hidden Markov Models (HMM) to differentiate expert surgical performances. Fuzzy classification methods have also been employed to assess bimanual skills and evaluate performance (Huang et al., 2005; Hajshirmohammadi and Payandeh, 2007). Jog et al. (2011) and Kerwin et al. (2012) have utilized decision tree algorithms to assess surgical simulator results. Jog et al also used Support Vector Machines (SVM) to compare their results with a decision tree algorithm. Loukas and Georgiou (2011) have presented 2 machine learning methods (SVM and HMM) for laparoscopic skills assessment of surgical trainees. Sewell et al. (2008) have proposed 3 different methods (HMM, Naïve Bayes, Logistic Regression) to assess performance of a surgical simulator. Richstone et al. (2010) have presented an Artificial Neural Network (ANN) method to assess surgical skill. Ershad et al. (2018) have used SVM to assess of robotic surgical performance. Wang and Fey (2019) have proposed a Convolutional Neural Network (CNN) for objective skill evaluation in robot-assisted surgery. Winkler-Schwartz et al. (2019) have assessed a series of machine learning algorisms used in virtual reality surgical studies to develop a series of best practice guidelines for these investigations.

A number of studies completed by our group have explored the use of machine learning and artificial neural networks to develop novel metrics and to classify surgical expertise utilizing both cranial and spine simulators. The study by Winkler-Schwartz et al. (2019) on virtual reality brain tumor resection used different machine learning algorithms (K-Nearest Neighbor, Naive Bayes, Discriminant Analysis and SVM) obtaining a 90% accuracy in classifying participants into 4 groups of expertise utilizing a NeuroVR simulator platform. Siyar et al. (2018) utilized 4 machine learning algorithms to classify “skilled” and “novice” participants employing a different brain tumor resection model obtaining an average of 90% accuracy. Bissonnette et al. (2019) utilized SVM and achieved a 97.6% ability to classify level of expertise utilizing a virtual reality spinal task on the NeuroVR simulator. Ledwos et al. (2020) validated an anterior cervical discectomy and fusion (ACDF) scenario on a novel spine simulator developed by OSSimTech and this scenario has been utilized in a series of new studies. Mirchi et al. (2020) have employed ANN for assessment of the virtual reality anterior cervical discectomy and fusion performance obtaining a test accuracy of 83.3% classifying 3 groups of participants based on level of expertise. The artificial neural network allowed for the differentiation of surgical expertise and provided insight into the relative importance of performance metrics. Mirchi et al. (2020) have proposed a virtual operative assistant as an explainable AI tool for simulation-based neurosurgical training employing SVM utilizing data from the study of Winkler-Schwartz et al. (2019) and they achieved a 92% accuracy in classifying “skilled” and “novice” groups.

In this study, we apply four different types of hybrid deep neural network algorithms called Fuzzy Clustering Neural Networks (FCNN-1, FCNN-2, FCNN-3, and FCNN-4) to assess if these networks can further improve classification of neurosurgical expertise using similar data as employed by Winkler-Schwartz et al. (2019).

2. FUZZY CLUSTERİNG NEURAL NETWORK

As seen in Fig. 1, Fuzzy Clustering Neural Network (FCNN) as hybrid learning method integrates both unsupervised Fuzzy C-means clustering and supervised artificial neural network. FCNN was defined and used by Karlik et al. (2002, 2003) and Ceylan et al. (2009, 2011) have improved this model and these methods have been named type-2 FCNN.

Clustering is the process of grouping into clusters in a dataset according to an appropriate similarity range for the degree of similarity so that the data points in a cluster are strong and weak in different clusters. Similarities of the data in cluster analysis are found by determining the distances from each other. For this, different distance methods are used. The most commonly used distance methods are Euclid, Manhattan and Minkowski Distance. The most popular clustering algorithms are k-means and Fuzzy C-Means. Fuzzy clustering points out different grades of membership from different clusters. As seen in Fig. 1, the fuzzy self-organizing layer is responsible for the clustering of the input data. The neural network is responsible for the classifier of obtained new training set according to target values. Here target values represent four levels of neurosurgical expertise. Karlik (2016) has proposed the positive effects of Fuzzy C-Mean Clustering on different supervised machine learning algorithms.

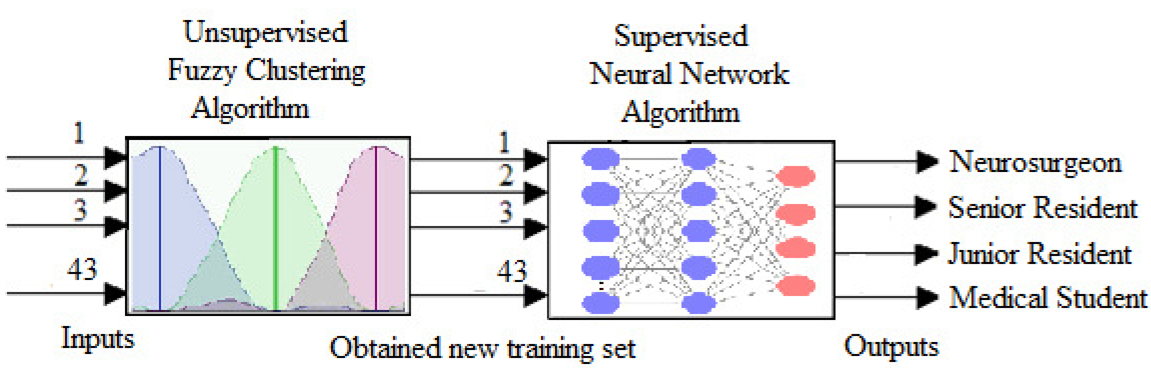

FIGURE 1: The block diagram of hybrid Fuzzy Clustering Neural Networks (FCNN).

The FCM algorithm calculates a suitable set of clusters in a limited number of cycles as an effective unsupervised algorithm (Karlik et al., 2009). It is important to designate different classes of membership from different clusters for FCM. Let X= {x1, x2,...,xn}, xi $\in R^p$ become the utilizing dataset. The FCM algorithm partitions the utilizing dataset into fuzzy groups (c) and computes cluster centers as V={v1,v2,.....,vc} $\subset R^p$ so that it can be minimized by

$$J=\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\, d_{ij}^{2} \quad (1)$$

where, $u_{ij} \in [0, 1]$ explains the degree of membership of the data point (xj) relating to the ith clustering, dij=‖vi-xj‖ is the Euclidean distance (Manhattan distance can be also used) between ith cluster center vi and jth data point xj, and ∊ [1, ∞) is an exponent of weighting which affects the fuzziness of the clusters. The equation (1) is minimized by

$$v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m} \quad (2)$$

and

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \quad (3)$$

subject to

$$\sum_{i=1}^{c} u_{ik} = 1 \; and \; 0 \le \sum_{k=1}^{n} u_{ik} \le n \quad (4)$$

Thus, reduction can be achieved using (uik). Simply, the FCM method is an iterative procedure through calculations of equations (2) and (3). The algoritm is working as:

Step 1: Set and initialize parameter values c, ∊ and uik.

Step 2: Compute Euclidean (or Manhattan) distances between cluster centers (vi), then if the distance is less than threshold cancel one of the releted centers by using Eq. (2).

Step 3: Calculate the distances using Euclidean distance (for type 1) or Manhattan distance (for type 2).

Step 4:Update cluster centers computing means in same cluster uik using Eq. (3).

Step 5:If max $\left\| u_{ik}^{NEW} - u_{ik}^{OLD} \right\| < \epsilon$, then stop. Otherwise, return to Step 2.

The output of FCM is fed to a supervised classifier as inputs as seen in Figure 1. The classifier algorithm has Backpropagation (BP) which consists of a multi layered perceptron (MLP) architecture. The BP searches for the mean squared error (MSE) in weight space using gradient descent technique. The combination of minimizing the error by updating the weights is considered as a solution of the learning for ANN. Since this technique requires calculation of the gradient of the error function at each step cycle, the continuity and differentiability of the error function must be guaranteed. It is possible to use different activation functions such as bipolar sigmoid, unipolar sigmoid, hyperbolic tangent, conic section function and radial basis function. In this study we found the best results for hyperbolic tangent and bipolar sigmoid activation function as proved by Karlik & Olgac with their FCNN software (Karlik and Olgac, 2011). Basically, backpropagation is a gradient descent method to minimize with error criteria, for each pattern. Computing of the gradient descent of weights for the whole training dataset is by

$$\Delta w_{ij}(n) = -\varepsilon * \frac{\partial E_p}{\partial w_{ij}} + \alpha * \Delta w_{ij}(n-1) \tag{5}$$

where n represents iteration cycle and are two positive coefficients called learning rate and momentum respectively. Momentum can accelerate training in very flat regions of the fault surface and suppresses weight oscillation in step valleys or ravines. But it requires propagation of all training dataset through the neural networks for computing. This can slow down the training process for big datasets.

Activation functions play an important role in the making of the deep neural networks (Karlik and Karlik, 2020; Bingham et al., 2020; Karlik and Karlik, 2021). However, the selection of activation functions also affects the network in terms of optimization and achieves better results. The different tasks can be facilitated by the correct choice of the transfer function in the output layer. So, the main novelty in this study is to use different transfer functions such unipolar sigmoid, bipolar sigmoid, radial based function (RBF), conic section function and hyperbolic tangent function. After using all these functions, we only added the results in the tables bipolar sigmoid and hyperbolic tangent functions which were giving better results than others. We did not add the other results because it gave bad results and not to make too many pages unnecessarily. To use the hyperbolic tangent (especially for the hidden layer) is right way mathematically. Because, the sigmoid (logistic) activation function will generate values close to 0.0 if the function argument is essentially negative. Thus, the output of this hidden neuron will be close to zero, and thus degradation the learning rate for all subsequent weights. Thus, it will nearly stop learning for complex dataset. The hyperbolic tangent function, will in the same situation generate a value close to -1.0, and thus will keep learning. In this paper, the four different combinations of hybrid fuzzy clustering neural network algorithms have been developed. And the results are comprised in table from 1 to 4.

3. MATERIALS AND METHODS

The goal of this study is to differentiate levels of surgical expertise in a virtual reality neurosurgical procedure known as a brain tumor resection by utilizing a series of performance metrics related to bimanual surgical skills and operative factors. A general block diagram of this application is given in Fig. 2. This system consists of data generation through the utilization of a virtul reality surgeical simuilation, data processing for metric generation, normalization and feature selection and lastly classification to asses surgical expertise.

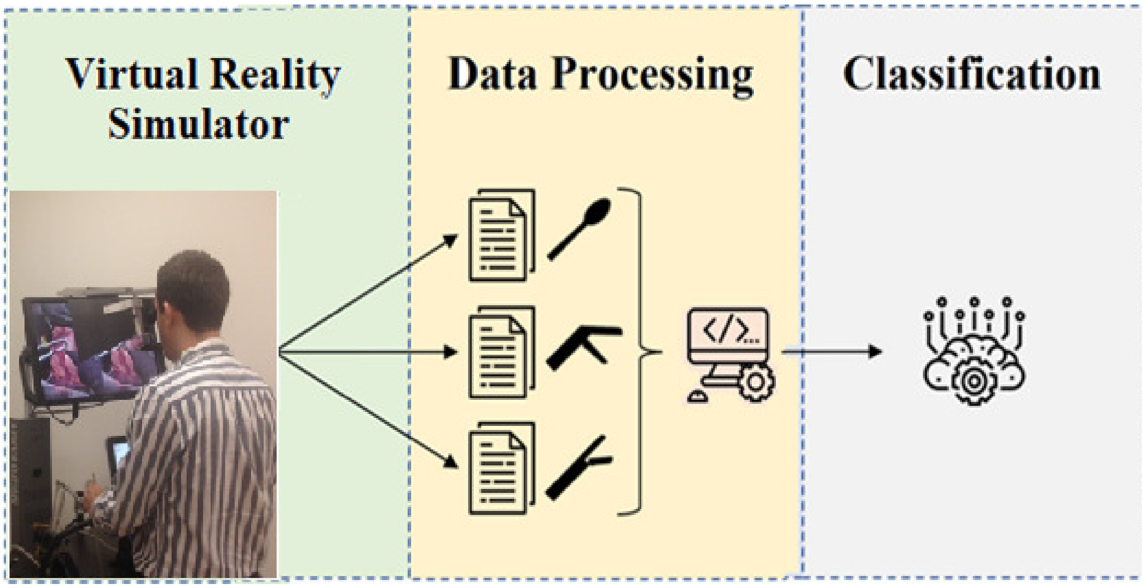

FIGURE 2: System of Assessment of Surgical Expertise in Virtual Reality Simulation.

3.1 Virtual Reality Neurosurgical Simulator

The NeuroVR simulator is a high fidelity platform which allows the analysis and interpretation of participant surgical performance carrying out simulated brain tumor resection procedures. The simulator consists of a microscope and handles where different surgical tools can be connected to simulate the performance of neurosurgeons during the surgical tasks (Delorme et al., 2012; Choudhury, 2013). The realism of the simulator is derived from its accurate anatomical display of the tissue and the haptic feedback. This gives the operator guidance as to the targeted tissue, how to manipulate these tissues and the amount of tissue that can be safely removed. A roadmap for creating complex virtual reality surgical simulation scenarios with a stepwise illustration of our team's subpial tumor resection model utilized in this study has been published previously (Sabbag, 2020).

The dataset was collected at a single time point from 50 participiants. These participants included 14 neurosurgeons and 4 neurosurgical fellows (experts), 10 senior residents, 10 junior residents and 12 medical students. Each individual performed the subpial tumor resection, a complex neurosurgical procedure, 5 times. All 250 simulated virtual reality neurosurgical tumor resection tasks were carried out as outlined by Winkler-Schwatz et al. (2019) and videos outlining the procedure are available in this publication.

This study was conducted at Neurosurgical Simulation and Artificial Intelligence Learning Centre at McGill University. The McGill University Health Centre Ethics Board, Neurosciences-Psychiatry approved this study and all participants signed an approved consent before the simulated subpial tumor resection. All procedures were in accordance with the ethical standards of the responsible committee on human exprimentation and with the Declaration of Helsinki.

3.2 Data Processing

The raw data recorded by the NeuroVR simulation underwent interpolation to fill missing data points and regularization before performance metric calculation. The performance metrics included the movement information such as speed, acceleration of instruments and the distance between the instruments as well as operative factors such as tumor and healthy tissue volumes removed, bleeding, and the number of attempts to stop bleeding. All features were calculated for three different operative conditions: during the whole task, while removing the tumor, and while suctioning blood. The average, median, and maximum values were obtained for each task. Finally, the performance measures for 5 repeated tasks were averaged for every participant, reaching 250 metrics for 50 participants. Using t-tests the number of metrics were reduced to 122 as outlined by Winkler-Schwartz et al. (2019)

Selected metrics with FCM were utilized as features of the dataset. In the other words, we presented an approach that the number of segments in the original training set was clustered by FCM. The selected dataset of metrics with FCM was normalized by using min-max and z score normalization methods to ensure optimal algorithm functioning. We obtained better results from min-max normalization technique.

3.3 Classification

The hybrid classifier used in this study is the integrated cascade connection of unsupervised fuzzy c-mean (FCM) and supervised neural network algorithms. First, FCM is used to obtain the cluster centers of whole dataset. This way, the center obtained from each cluster provides a new attribute to describe the metrics of feature sets. The 122 sample feature set corresponding to each of the 4 different groups are clustered using the FCM algorithm before the applying ANN. The clustered data is applied as input into the ANN which is trained with the back-propagation algorithm. Data was clustered with different features as 2, 4, 6, 8 and 10. The best result was obtained for 8. Then selected data was normalized using min-max normalization method. The architectures that are trained by four different hybrid methods called as FCNN-1, FCNN-2, FCNN-3 and FCNN-4. FCNN-1 consist of Euclidean distance for FCM algorithm and bipolar sigmoid activation function for BP algorithm. Similarly, FCNN-2 has Manhattan distance and bipolar sigmoid function; FCNN-3 has Euclidian distance and hyperbolic tangent function, FCNN-3 has Manhattan distance and hyperbolic tangent function respectively. The backpropagation algorithm (BP) was utilized for training of neural networks employed different types of transfer functions. A screenshot of the proposed FCNN software is shown in Fig. 3.

Here, the structure of MLP 122:122:4 which means number of nodes of input layer, nodes of hidden layer and nodes of output layer were utilized as 122, 122 and 4, respectively. The node number for the hidden layer were determined as 122 via experimentation for all FCNN architectures. Each output represents each class of neurosurgical expertise as neurosurgeons (representing practicing neurosurgeons and neurosurgical fellows), senior residents, junior residents and medical students. Optimum learning constant and momentum coefficient were selected as 0.9 and 0.7 in training via experimentations according error.

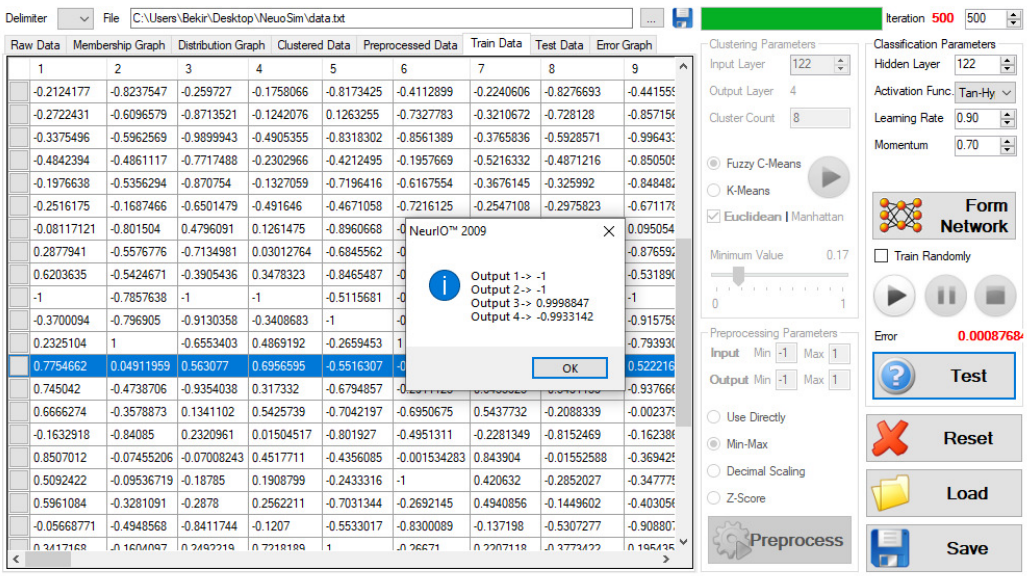

FIGURE 3: Graphical Interfaces of proposed FCNN.

4. RESULTS OF PROPOSED ALGORITHMS

In this study, 4 different architectures of Fuzzy Clustering Neural Networks (FCNN) have been utilized for the assessment of surgical expertise in virtual reality simulation. A cross-validation method was used as a resampling procedure to evaluate the simplest choice for split tests on a limited data sample. The test results are outlined in Table-1 to Table-4 respectively. If the test output of the FCNN is positive that means the output = +1. In the case of surgical simulation, this corresponds to a skilled participant. However, if the test output of the FCNN is negative, the output = -1 this represents a less skilled participant. For these tables all results were found for 500 iterations. If we increase number of iterations, the accuracy of all models will also be increased. The following two Tables were obtained for bipolar sigmoid activation functions. We have also used the other activation functions such as unipolar sigmoid, conic section and radial basis functions (RBF) for the same FCNN models but we did not find better results than bipolar sigmoid and hyperbolic tangent functions. Therefore, these results were not added here.

TABLE 1: FCNN-1 (Euclidean and bipolar sigmoid) results for 500 iterations.

	Neurosurgeons	**Senior Residents**	**Junior Residents**	**Medical Students**
Neurosurgeons	12 (85.71%)	2 (14.29%)	0 (0%)	0 (0%)
Senior Residents	1 (7.1%)	13 (92.9%)	0 (0%)	0 (0%)
Junior Residents	0 (0%)	1 (10%)	9 (90%)	0 (0%)
Medical Students	0 (0%)	0 (0%)	0 (0%)	12 (100%)

Error: 0.011992

TABLE 2: FCNN-2 (Manhattan and bipolar sigmoid) results for 500 iterations.

	Neurosurgeons	Senior Residents	Junior Residents	Medical Students
Neurosurgeons	11 (78.57%)	2 (14.29%)	1 (7.14%)	0 (0%)
Senior Residents	0 (0%)	14 (100%)	0 (0%)	0 (0%)
Junior Residents	0 (0%)	1 (10%)	9 (90%)	0 (0%)
Medical Students	0 (0%)	0 (0%)	0 (0%)	12 (100%)

Error: 0.011687

These results were found for 500 iterations for both (Euclidian and Manhattan) distance types and the activation function were bipolar sigmoid for neural networks. In this method, useful min-max normalization technique (between 0 and 1 interval) was used. We also used other normalization techniques such as z-score and decimal scaling, but we did not find better accuracy (or less error) than min-max technique. This type of FCNN had an accuracy of 92% (46 of 50). But when we increased the number of iterations from 500 to 1500, accuracy was increased and reached 100% (50 of 50).

Similarly, we found better results by using hyperbolic tangent (tanh) activation function for 500 iterations as seen in Table 3 and Table 4. In these structures, the min-max normalization technique (between -1 and 1 interval) were used. For all FCNN types, optimum learning rate and momentum coefficients were 0.9 and 0.7 respectively.

TABLE 3: FCNN-3 (Euclidean and tanh) results for 500 iterations.

	Neurosurgeons	Senior Residents	Junior Residents	Medical Students
Neurosurgeons	14 (100%)	0 (0%)	0 (0%)	0 (0%)
Senior Residents	0 (0%)	14 (100%)	0 (0%)	0 (0%)
Junior Residents	0 (0%)	0 (0%)	10 (100%)	0 (0%)
Medical Students	0 (0%)	0 (0%)	0 (0%)	12 (100%)

Error: 0.0016594

TABLE 4: FCNN-4 (Manhattan and tanh) results for 500 iterations.

	Neurosurgeons	Senior Residents	Junior Residents	Medical Students
Neurosurgeons	14 (100%)	0 (0%)	0 (0%)	0 (0%)
Senior Residents	0 (0%)	14 (100%)	0 (0%)	0 (0%)
Junior Residents	0 (0%)	0 (0%)	10 (100%)	0 (0%)
Medical Students	0 (0%)	0 (0%)	0 (0%)	12 (100%)

Error: 0.0016588

5. CONCLUSION AND DISCUSSION

Improved classification accuracy results when a classifier performs more efficently in the classification (Akyol et al., 2016). To improve the accuracy of the classifier, a pretreatment step may be required that can eliminate irrelevant data, noise and unnecessary features from sensors or measuring devices. This irrelevant data problem is frequently seen in medical data which consists of various features. Cascade connection of unsupervised clustering and supervised classifier is one of best hybrid machine learning models for medical data (Ceylan et al., 2014; Esme and Karlik, 2016; Mehta et al., 2018). Hybrid classifier also has the potential to combine strengths and to overcome shortcomings.

The aim of this study was to differentiate with high accuracy, levels of neurosurgical expertise by an effective hybrid deep neural network algorithm called Fuzzy Clustering Neural Networks in a virtual reality surgical brain tumor resection task. For this purpose, four hybrid classification models (FCNN1, FCNN2, FCNN3 and FCNN4) were used and compared with conventional machine learning algorithms. Table 5 outlines results comparing our previous studies to those of our present investigation.

TABLE 5: Results of different machine learning algorithms for the 4 groups.

Used Methods	Accuracy (%)			
	Neurosurgeons	Senior Residents	Junior Residents	Medical Students
k-NN	85.7	92.9	90.0	91.7
LDA	85.7	71.4	80.0	75.0
Naïve Bayes	92.9	78.6	70.0	91.7

SVM	78.6	85.7	60.0	75.0
FCNN-1	85.7	92.9	90.0	**100.0**
FCNN-2	78.6	**100.0**	90.0	**100.0**
FCNN-3	**100.0**	**100.0**	**100.0**	**100.0**
FCNN-4	**100.0**	**100.0**	**100.0**	**100.0**

The results demonstrate that the proposed hybrid algorithms employed have improved the accuracy of classification when compared to the previous studies utilizing the same dataset. In the previous studies (Winkler-Schwartz et al. (2019), K-NN achieved the best accuracy at 90% with 2 neurosurgeons, 2 senior residents, and 1 medical student misclassified (Table 5). Both FCNN-1 and FCNN-2 models obtained slightly improved accuracy of 92%. By using the proposed FCNN-3 and FCNN-4 models, we achieved a further improvement in classification accuracy of 100% correctly classifying the level of expertise of 50 out of 50 participants.

There may be a number of reasons for the improved level of expertise classification utilizing the FCNN-3 and FCNN-4 models. First, the study of Winkler-Schwartz et al. (2019), used less metrics then were employed utilizing the FCNN-3 and FCNN-4 models. Using less metrics and the resulting decrease in computational complexity may be related to the reduced performance of the machine learning algorisms used. Second, our previous studies used the z-scores normalization while the present study used the min-max normalization method which may have improved the ability of the FCNN-3 and FCNN-4 models to differentiate levels of expertise.

The dataset used in these studies is not consider big data and the accuracy of classification is excellent using the FCNN-3 and FCNN-4 models. We are planning to significantly expand the data size by increasing the number of surgical residents performing simulated brain tumor resections using the NeuroVR platform. For much larger datasets, it may be necessary to use our hybrid methods in combination with deep learning methods to obtain optimal results (De Campos Souza, 2020).

6. REFERENCES

Akyol, A., Hacibeyoglu, M., Karlik, B. (2016). Design of Multilevel Hybrid Classifier with Variant Feature Sets for Intrusion Detection System. *IEICE Transactions on Information and Systems*, E99-D, 7, 1810-1821. https://doi: 10.1587/transinf.2015EDP7357.

Alotaibi, F. E., AlZhrani, G. A., Mullah, M. A., Sabbagh, A. J., Azarnoush, H., Winkler-Schwartz, A., et al. (2015). Assessing bimanual performance in brain tumor resection with NeuroTouch, a virtual reality simulator. *Neurosurgery*, *11*(2), 89-98; discussion 98, https://doi:10.1227/NEU.0000000000000631.

Bissonnette, V., Mirchi, N., Ledwos, N., Alsidieri, G., Winkler-Schwartz, A., Del Maestro, R.F. (2019). Artificial ıntelligence distinguishes surgical training levels in a virtual reality spinal task. *Journal of Bone and Joint Surgery*, vol. 101(23) e127. https://doi:10.2106/jbjs.18.01197PMID: 31800431.

Bingham, G., Macke, W., Miikkulainen, R.A. (2020). Evolutionary optimization of deep learning activation functions. GECCO'20: Proceedings of the 2020 Genetic and Evolutionary Computation Conference, July 8 – 12, 2020, Cancún Mexico, 289–296. https://doi.org/10.1145/3377930.3389841.

Ceylan, R., Ozbay, Y., Karlik, B. (2009). Classification of ECG arrhythmias using type-2 fuzzy clustering neural network. 14th National Biomedical Engineering Meeting, Balcova, Izmir, 2009. 1-4, https://doi: 10.1109/BIYOMUT.2009.5130250.

Ceylan, R., Ozbay, Y., Karlik, B. (2011). Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ECG classifier. *Expert Systems with Applications*, *38*, 1004-1010. https://doi: 10.1016/j.eswa.2010.07.118.

Ceylan, R., Ozbay, Y., Karlik, B. (2014). Comparison of type-2 fuzzy clustering-based cascade classifier models for ECG arrhythmias. *Biomedical Engineering Applications Basis and Communications*, *26*(06):1450075. https://doi:10.4015/S1016237214500756.

Choudhury, N., Gélinas-Phaneuf, N., Delorme, S., Del Maestro, R. (2013). Fundamentals of neurosurgery: virtual reality tasks for training and evaluation of technical skills. *World Neurosurg*., *80*(5), 9-19. https://doi: 10.1016/j.wneu.2012.08.022.

De Campos Souza, P.V. (2020). Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature, *Applied Soft Computing*, 92, 106275. https://doi.org/10.1016/j.asoc.2020.106275.

Delorme, S., et al. (2012). NeuroTouch: A physics-based virtual simulator for cranial microneurosurgery training. *Neurosurgery*, *71*(1), 32-42. https://doi:10.1227/NEU.0b013e318249c744.

Esme, E. and Karlik, B. (2016). Fuzzy c-means based support vector machines classifier for perfume recognition. *Applied soft computing*, *46*, 452-458. https://doi.org/10.1016/j.asoc.2016.05.030.

Ershad, M., Rege, R., Fey A. M. (2018). Meaningful Assessment of Robotic Surgical Style using the Wisdom of Crowds. *Int J Comput Assist Radiol Surg*., *13*(7), 1037-1048. https://doi:10.1007/s11548-018-1738-2.

Fried, G. M., Feldman, L. S., Vassiliou, M. C., Fraser, S. A., Stanbridge, D., Ghitulescu, G., et al. (2004). Proving the value of simulation in laparoscopic surgery. *Annals of Surgery*, *240*(3), 518-528. https://doi:10.1097/01.sla.0000136941.46529.56.

Gelinas-Phaneuf, N., and Del Maestro, R. F. (2013). Surgical expertise in neurosurgery: integrating theory into practice. *Neurosurgery*, *73*, 30-38. https://doi:10.1227/NEU.0000000000000115.

Hajshirmohammadi, I. and Payandeh, S. (2007). Fuzzy set theory for performance evaluation in a surgical simulator. Presence: *Teleoperators Virtual Environ*., *16*(6), pp.603-622.

Huang J, Payandeh S, Doris P, Hajshirmohammadi I. (2005). Fuzzy classification: towards evaluating performance on a surgical simulator. *Stud Health Technol Inform*., *111*, 194-200. http://ebooks.iospress.nl/publication/10064. Accessed August 23, 2018.

Jog, A., Itkowitz, B., Liu, M., et al. (2011). Towards integrating task information in skills assessment for dexterous tasks in surgery and simulation. IEEE Inter. Conf. on Robotics and Automation. 5273-5278. https://doi:10.1109/ICRA.2011.5979967.

Karlık, B., Tokhi, O., Alcı, M. (2002). A novel technique for classification of myoelectric signals for prosthesis, 15th Triennial World Congress, (IFAC'02), Barcelona, Spain 21 - 26 July 2002. https://doi: 10.3182/20020721-6-ES-1901.00980.

Karlık, B., Tokhi, O., Alcı, M. (2003). A fuzzy clustering neural network architecture for multi-function upper-limb prosthesis. *IEEE Trans. Biomedical Eng*., *50*(11), 1255–1261. https://doi: 10.1109/TBME.2003.818469.

Karlik, B. (2016). The positive effects of fuzzy c-mean clustering on supervised learning classifiers. *Inter Journal of Artificial Intelligence and Expert Systems*, *7*(1), 1-8. https://doi:10.1.1.740.500&rep=rep1&type=pdf.

Karlik, B., Korürek, M., Koçyiğit, Y. (2009). Differentiating types of muscle movements using wavelet based fuzzy clustering neural network. *Expert Systems*, *26*(1), 49–59. doi: https://doi.org/10.1111/j.1468-0394.2008.00496.x.

Karlik, B. and Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Inter Journal of Artificial Intelligence and Expert Systems*, *4*, 111-122. https://doi:10.1.1.740.9413&rep=rep1&type=pdf.

Karlik, M. and Karlik, B. (2020). Prediction of student's performance with deep neural networks. Inter Journal of Artificial Intelligence and Expert Systems, *9*(2), 39-47.

Karlik, M. and Karlik, B., (2021). Second language learning with affective factors and deep neural networks methods. *Canadian Journal of Language and Literature Studies*, *1*(3), 26-43. https://doi.org/10.53103/cjlls.v1i3.20.

Kerwin T, Wiet G, Stredney D, Shen H.W. (2012). Automatic scoring of virtual mastoidectomies using expert examples. *Int Journal of Comput Assist Radiol Surg*., *7*(1), 1-11. https://doi:10.1007/s11548-011-0566-4.

Liang, H., Shi, M.Y. (2011). Surgical skill evaluation model for virtual surgical training. *ApplMech Mater*., *40-41*, 812-819. https://doi:10.4028/www.scientific.net/AMM.40-41.812.

Ledwos, N., Mirchi, N., Bissonnette, V., Winkler-Schwartz, A., Yilmaz, R., Del Maestro, R.F. (2021). Virtual Reality Anterior Cervical Discectomy and Fusion Simulation on the Novel Sim-Ortho Platform: Validation Studies. *Operative Neurosurgery, 20*(1), 74–82. https://doi.org/10.1093/ons/opaa269.

Loukas C, Georgiou E. (2011). Multivariate autoregressive modeling of hand kinematics for laparoscopic skills assessment of surgical trainees. *IEEE Trans Biomed Eng*., *58*(11), 3289-3297. https://doi:10.1109/TBME.2011.2167324.

Megali, G., Sinigaglia, S., Tonet, O., and Dario, P. (2006). Modelling and evaluation of surgical performance using hidden Markov models. *IEEE Trans Biomed Eng*, *53*(10), 1911-1919. https://doi:10.1109/TBME.2006.881784.

Mehta, R., Bath, S.K., Sharma, A.K. (2018). An analysis of hybrid layered classification algorithms for object recognition. *IOSR Journal of Computer Engineering*, *20*(1), 57-64. https://doi: 10.9790/0661-2001035764.

Mirchi, N., Bissonnette, V., Ledwos, N., Winkler-Schwartz, A., Yilmaz, R., Karlik, B., Del Maestro, R.F. (2020). Artificial Neural Networks to Assess Virtual Reality Anterior Cervical Discectomy Performance. *Operative Neurosurgery*, *19*(1), 65–75. https://doi.org/10.1093/ons/opz359.

Mirchi, N., Bissonnette, V., Yilmaz, R., Winkler-Schwartz, A., Del Maestro, R.F. (2020). The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation based training in surgery and medicine. *PLoS ONE*, *15*(2), e0229596. https://doi.org/10.1371/journal.pone.0229596.

Moorthy, K., Munz, Y., Sarker, S. K., Darzi, A. (2003). Objective assessment of technical skills in surgery. *British Medical Journal*, *327*(7422), 1032-1037.

Rhienmora P, Haddawy P, Suebnukarn S, Dailey M.N. (2011). Intelligent dental training simulator with objective skill assessment and feedback. *Artificial Intell Medicine*, *52*(2), 115-121. https://doi:10.1016/j.artmed.2011.04.003.

Richstone. L., Schwartz, M. J., Seideman, C, Cadeddu, J., Marshall, S., Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Ann Surg*., *252*(1), 177-182. https://doi:10.1097/SLA.0b013e3181e464fb.

Sabbag, A.J., Bajunaid, H.M., Alarifi, N., Winkler-Schwartz, A., Alsideiri, G., Al-Zhrani, G., Alotaibi, F.E., Bugdadi, A., Laroche, D., Del Maestro, R.F. (2020). Roadmap for developing complex virtual reality simulation scenarios: subpial neurosurgical tumor resection model. *World Neurosurgery*, *139*, e220-e229. https://doi.org/10.1016/j.wneu.2020.03.187.

Sewell, C., Morris, D., Blevins, N.H., et al., Providing metrics and performance feed-back in a surgical simulator. (2008). *Comput Aided Surg*., *13*(2), 63-81. https://doi:10.1080/10929080801957712.

Siyar, S. Azarnoush, H., Rashidi, S., Winkler-Schwartz, A., Bissonnette, V., Ponnudurai, N., Del Maestro, R.F. (2018). Machine learning distinguishes neurosurgical skill levels in a virtual reality tumor resection task.https://arxiv.org/ftp/arxiv/papers/1811/1811.08159.pdf.

Wang, Z. and Fey, A.M. (2019). Convolutional Neural Network (CNN) for objective skill evaluation in robot-assisted surgery. https://arxiv.org/abs/1806.05796.

Winkler-Schwartz, A., Bissonnette, V., Mirchi, N., Ponnudurai, N., Yilmaz, R., Ledwos, N., Siyar, S., Azarnoush, H., Karlik, B., Del Maestro, R.F. (2019). Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. *Journal of surgical education*, *76*(6), 1681-90.

Winkler-Schwartz, A., Yilmaz, R., Mirchi, N., Bissonnette, V., Ledwos, N., Siyar, S., et al. (2019). Assessment of machine learning identification of surgical operative factors associated with surgical expertise in virtual reality simulation. *JAMA Network Open*., *2*(8), e198363–e198363, https://doi.org/10.1001/jamanetworkopen. PMID: 31373651.