Analyze Informant-Based Questionnaire for The Early Diagnosis of Senile Dementia Using Deep Learning

FUBAO ZHU[1], XIAONAN LI[1], DANIEL MCGONIGLE[2], HAIPENG TANG[2], ZHUO HE[3], CHAOYANG ZHANG[2], GUANG-UEI HUNG[4], PAI-YI CHIU[5], AND WEIHUA ZHOU [ID][3]

[1]School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China
[2]School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Long Beach, MS 39560, USA
[3]College of Computing, Michigan Technological University, Houghton, MI 49931, USA
[4]Department of Nuclear Medicine, Chang Bing Show Chwan Memorial Hospital, Changhua 505, Taiwan
[5]Department of Neurology, Show Chwan Memorial Hospital, Changhua 500, Taiwan

CORRESPONDING AUTHORS: P.-Y. CHIU (paiyibox@gmail.com) AND W. ZHOU (whzhou@mtu.edu)

ABSTRACT Objective: This paper proposes a multiclass deep learning method for the classification of dementia using an informant-based questionnaire. Methods: A deep neural network classification model based on Keras framework is proposed in this paper. To evaluate the advantages of our proposed method, we compared the performance of our model with industry-standard machine learning approaches. We enrolled 6,701 individuals, which were randomly divided into training data sets (6030 participants) and test data sets (671 participants). We evaluated each diagnostic model in the test set using accuracy, precision, recall, and F1-Score. Results: Compared with the seven conventional machine learning algorithms, the DNN showed higher stability and achieved the best accuracy with 0.88, which also showed good results for identifying normal (F1-score = 0.88), mild cognitive impairment (MCI) (F1-score = 0.87), very mild dementia (VMD) (F1-score = 0.77) and Severe dementia (F1-score = 0.94). Conclusion: The deep neural network (DNN) classification model can effectively help doctors accurately screen patients who have normal cognitive function, mild cognitive impairment (MCI), very mild dementia (VMD), mild dementia (Mild), moderate dementia (Moderate), and severe dementia (Severe).

INDEX TERMS Dementia, information gain, deep neural network, machine learning.

I. INTRODUCTION

Dementia characterized by cognitive and intellectual impairment is a kind of neurodegenerative diseases [1]. Dementia diagnosis is a critical issue since it affects 47.5 million people worldwide according to World Health Organization [2]. Currently, cognitive and/or memory disorders are the primary metrics [3] used to identify whether individual suffers the dementia or not. To assess the cognitive status of a patient the neuropsychological tests are commonly used in clinical diagnosis. Nevertheless, it is time-consuming for the manual diagnosis of cognitive impairment by using neuropsychological tests. Moreover, the efficiency and accuracy of the diagnosis are determined by the professional level of the practitioner. In some remote areas lacking professional personnel, it will be a much more difficult task for classification and the early diagnosis of dementia.

In recent years, deep neural network (DNN) has garnered clinical interests in cognitive diagnostic applications due to its advantages in efficient classification. Moreover, many existing methods have been proposed, where some of them [4]–[7] combine DNN with neuroimaging markers, while other methods [8]–[10] combine DNN with neuropsychological assessments. Jain et al. [5] proposed a transfer learning approach for accurately classifying brain sMRI slices amongst 3 different classes: Alzheimer's disease (AD), cognitively normal(CN) and mild cognitive impairment (MCI). For the validation set, the accuracy of the three-way classification using their method was 95.73%. However, they only

analyzed the T1-weighted sMRI data of 150 subjects/patients. Lu *et al.* [6] proposed a novel deep-learning-based framework to discriminate individuals with AD utilizing a multimodal and multiscale deep neural network. They obtained an accuracy of 82.4% in identifying the individuals with MCI. They analyzed both a T1-weighted MRI scan and FDG-PET image data of 1242 subjects/patients. Orimaye *et al.* [8] proposed a method that combined deep neural network and deep language models (D2NNLM) for classifying the disease. The experimental results showed that the model could accurately predict MCI and AD type dementia on a very sparse clinical language dataset. Themistocleous et al. [10] provided an automated deep learning method using DNN architectures that identified individuals with MCI from healthy controls. However, there are still limitations in the current studies. First, the amount of data used in the methods introduced above was not sufficient. This limitation may lead to a decrease in the reliability of experimental results. Second, these methods focused more on binary classification problem. Yet, as dementia is currently irreversible and incurable, multi-class classification for different dementia stages is actually of much more clinical interest.

Hence, we propose a DNN multi-class classification model to assist the preliminary diagnosis of normal, mild cognitive impairment (MCI), very mild dementia (VMD), mild dementia (Mild), moderate dementia (Moderate), and severe dementia (Severe) using informant-based questionnaire. In this paper, 6,701 individuals are enrolled, which allow us to have a larger size of samples to train our model.

II. MATERIALS AND METHODS

There are two major steps in the proposed framework: (1) feature selection: using feature selection algorithms to optimize or even reduce the number of neuropsychological tests; (2) classification: training a deep neural network to classify the participants into normal cognitive function, MCI, VMD, Mild, Moderate, and Severe categories.

A. PATIENT SAMPLE COLLECTION

In this work, the study used data collected from the three centers of the Show Chwan Healthcare System. The data selected from the register-based database of the Show Chwan Health System were analyzed anonymously with the informed consent from all participants, and the study was designed retrospectively in accordance with relevant guidelines and regulations. The project was reviewed by the Medical Research Ethics Committee of Show Chwan Memorial Hospital, and the study was approved by the Data Inspectorate.

The data for the study consisted of samples of clinical and neuropsychological assessment obtained from 6701 patients. For detailed neuropsychological tests, we assessed the history of cognitive status and objective assessments including the Clinical Dementia Ratings (CDR), Mini Mental Status Examination (MMSE), Cognitive Abilities Screening Instrument (CASI) and Montreal Cognitive Assessment (MoCA) performed to evaluate memory, executive function,

orientation, visual-spatial ability, and language function [11]. Along with the current scales such as CDR, MMSE, CASI, MoCA, we used a newly designed Informant-based questionnaire named HAICDDS which is applied in dementia registration in a health system with 9 regional hospitals in Taiwan. Clinical application of the HAICDDS had been published in journals [11]–[13] or conferences [14], [15]. The CDR determined the severity of dementia. Experienced neurologists evaluated the participants based on their clinical symptoms and reviews of medical/medication history, neuropsychological test results, and then classified the participants into six diagnostic groups: normal (535 participants), MCI (1687 participants), VMD (678 participants), Mild (1812 participants), Moderate (1309 participants), and Severe (680 participants). The six diagnostic groups were defined using the CDR staging. Among CDR 0.5, participants without significantly impaired activities of daily living were divided as CDR 0.5 MCI and those with significantly impaired activities of daily living were divided as CDR 0.5 VMD. Therefore, the 6 groups were CDR 0, CDR 0.5 MCI, CDR 0.5 VMD, CDR 1, CDR 2, and CDR 3. The operational diagnosis of a significant interfere with ADL is the IADL total score <7.

we randomly split the data with the ratio of 9:1, of which 90% are training data sets (6030 participants) and 10% are test data sets (671 participants)[16]. In order to estimate the generalization error, this procedure was repeated 10 times independently to avoid any deviation caused by randomly partitioning data sets. The average accuracy and F1-score were calculated for performance analysis. We finally obtained 10 training-test of different training set (6030 participants) and test set (671 participants). We also repeated the independent training-test procedure more than 10 times (k = 10) but the results were similar, so only the results with k = 10 were reported in the manuscript.

B. FEATURE SELECTION

Neuropsychologists selected 50 items from neuropsychological tests to form an optimal questionnaire for screening patients with varying severities of dementia. While the proposed algorithm still works with the entirety of those features, utilizing feature selection lowers the computational requirements and improves interpretability as we are able to see more clearly which items are directly correlated with a patient's mental condition [11]. In order to retain the features with higher prediction performance, provide faster and more cost-effective predictors, reduce the curse of dimensionality problem and the possibility of overfitting during the training phase, we used information gain feature selection algorithms to rank the importance score of all 50 features and then the low ranking features were filtered out. We discard features with a lower score one by one, and input the remaining features into the DNN model to observe the change of the classification accuracy in order to identify the feature set with a smaller number of features but only a minor drop-off of classification accuracy.

Information gain is an information theory method widely used in data mining. The information gain measures how much information the feature can provide the classification model. If a feature has a larger information gain value for a class, the feature contains more classification information for that class. We used the information gain algorithm provided by Weka, which is an open source machine learning and data mining software based on the Java environment.

C. OVERVIEW OF METHODS

We proposed a DNN classification model based on the Keras framework. In order to study the performance of the DNN model for discriminating normal, MCI, VMD, Mild, Moderate, and Severe, we compared its results with others well-known classification models (MLP, GCForest, random forest, AdaBoost, LogitBoost, Naïve Bayes and SVM). First, we tested the performance of the model respectively using 50 features selected by the Neuropsychologists and the top 44 features selected by information gain score. Then, for testing the stability of the classification model, we tested 10 training-test runs separately. Finally, we evaluated the accuracy of model classification by measuring the average accuracy of the 10 training-test runs and evaluated the classification performance of each sub-category by using the 10 training-test runs with the lowest F1-score. TP is the number of positive samples predicted by the classifier in the number of true positive samples, FP is the number of positive samples predicted by the classifier in the number of true negative samples, TN is the number of negative samples predicted by the classifier in the number of true negative samples, FN is the number of true positive samples predicted by the classifier as negative samples. The accuracy is the evaluation of the correct rate of the classifier as a whole. It is defined as Accuracy = (TP+TN)/(TP+FN+FP+TN). Generally speaking, the higher the accuracy, the better the classifier. F1-Score is a kind of statistic, which is also called F-measure. F1-Score is the weighted harmonic average of Precision and Recall(Sensitivity). It is defined as F1-Score = 2TP/(2TP+FP+FN) and carries a range of 0 to 1, with higher scores indicating a more robust classification model. It is a commonly used evaluation criterion in the field of IR (Information Retrieval). It is often used to evaluate the quality of classification models. Precision, also called Positive Predictive Value in clinical settings, refers to how many of the samples that the model is positive are true positive samples, which is defined as Precision = TP/(TP+FP). Recall, also called Sensitivity in clinical settings, refers to how many positive samples are classified as positive by the model, which is defined as Recall = TP/(TP+FN).

1) DNN

DNN is a multi-hidden layer feedforward neural network, which has a total of L+1 layers, the 0th layer is the input layer, the 1st to L-1 layers are hidden layers, the Lth layer is the output layer. The nodes of adjacent layers are connected by links and the weights of all links form a feedforward

weight matrix. As shown in Eqs.(1)-(2), suppose there are neurons in the lth layer, and the input vector of these neurons is $a^{(l)}$, the output vector is $z^{(l)}$. At the same time, we distinguish the final output of DNN on the output of the hidden layer by u = $y^{(L)}$. Given the characteristic x of a training sample, there is $a^{(0)} = z^{(0)} = x$.

$$a^{(l)} = W^{(l)} * a^{(l-1)} + b^{(l)}, \quad l = 1, 2, \ldots L \quad (1)$$
$$z^{(l)} = h(a^{(l)}) \quad (2)$$

where $W^{(l)}$ is the weight matrix of the l-1 layer to the lth layer, b is the offset vector of the lth layer, h() is the activation function of the lth layer.

As a feedforward neural network, given an input vector, DNN can get an output vector immediately, that is to say, the output of DNN only depends on the current input, so DNN is suitable for pattern classification problem. This paper adjusts the network parameters of the whole network through supervised training. After repeated extensive training, we got relatively optimal hyperparameters in DNN. The constructed DNN model consists three hidden layers, the first layer uses the relu (rectified linear unit) activation function, the second layer uses the tanh [17] activation function, the third layer uses the softmax activation function. The epoch is set to 40, the dropout rate is set to 0.2, the batch size is set to 32, the learning rate is set to 0.004, and the number of neurons is set to 20 in each layer. The specific structure of the DNN model is shown in the figure 1 below. where *l* is the number of layers, x is the input feature, b is the offset vector of the lth layer, $W^{(l)}$ is the weight matrix of the l-1 layer to the lth layer, $a^{(l)}$ is the input vector of the lth layer, h() is the activation function of the lth layer, $z^{(l)}$ is the input vector of the next layer, f ( ) is the output activation function and y is the output vector.

2) OTHER INVESTIGATED CLASSIFICATION MODELS

We investigated other commonly used classification models (MLP, GCForest, random forest, AdaBoost, LogitBoost, Naïve Bayes and SVM) in Python toolbox[16], which is a set of freeware academic software packages. The following briefly describes the basic principles of the classification models and the further details can refer to cited literatures.

An MLP can be seen as a directed graph, consisting of multiple node layers, each layer connected to the next layer. In addition to input nodes, each node is a neuron (or processing unit) with a non-linear activation function0. Unlike the DNN model, the output activation function is not used here. GCForest is a model of a deep forest, which is mainly divided into two parts, multi-grained scanning, and cascade forest structure. GCForest performs well in small sample data. Random forest is an algorithm that integrates multiple trees by the idea of ensemble learning. Its basic unit is a decision tree, which is a subclass of ensemble learning. It depends on the voting choice of a decision tree to determine the final classification results. In the basic Adaboost algorithm, each weak classifier has the right to weight, and the
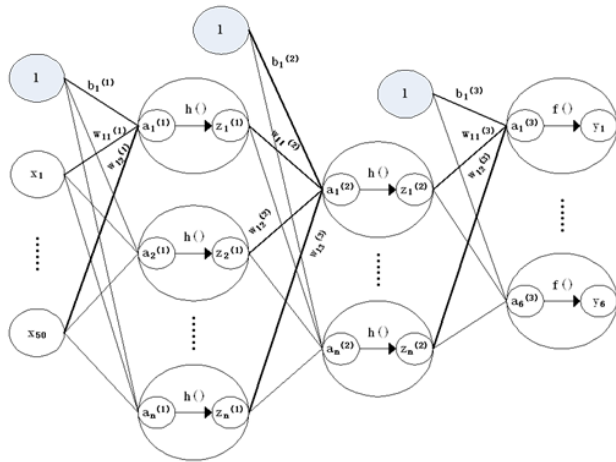
FIGURE 1. DNN model structure.

FIGURE 2. Classification accuracy with decreasing number of features.

FIGURE 3. Features ranked according to their information gain scores.

(a) 50_feature

(b) Top44_feature

FIGURE 4. Performance of accuracy for each of the 10 rounds.

weighted sum of the weak classifier prediction results forms the final prediction result. In training, training samples have also weight, which dynamically adjusts during the training process. The samples that are misclassified by the previous weak classifier will increase the weight, so the algorithm will focus on the difficult samples. The Logitboost algorithm is a discriminant classification algorithm based on machine learning. LogitBoost belongs to the AdaBoost system. The LogitBoost structure is similar in general, but its loss function uses the maximum logarithmic likelihood function. The basic method of Naïve Bayes is to calculate the probability that the current feature samples belong to a certain classification based on the statistical data and the conditional probability formula, and select the maximum probability classification0. Support Vector Machine (SVM) has achieved the best performance in many classification problems. The kernel function subtly transforms the linear indivisible problem into a linear separable problem, and has very good generalization performance.

III. RESULTS
A. FEATURE ANALYSES
We find that the classification accuracy decreases with the decrease of the number of features. When the number of features decreases to 44 features, the classification accuracy dropped down. So we discarded the corresponding features by setting the threshold of the information gain score to 0.16.

Figure 2 shows the trend of classification accuracy by our DNN model as the features with lower scores are discarded one by one. The classification accuracy is the average of ten experiments. With a decreasing number of features, the classification accuracy decreases. After reducing to 44 features, the subsequent classification accuracy has declined by a certain extent.
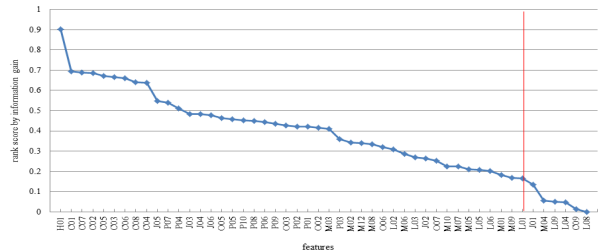
Figure 3 shows the features ranked in descending significance with respect to the information gain scores. The cut-off is shown reducing the number of features to the 88% with setting the threshold of the information gain score to 0.16,
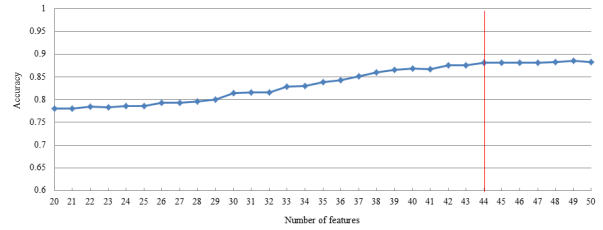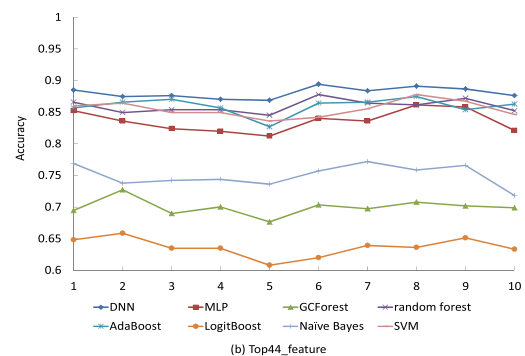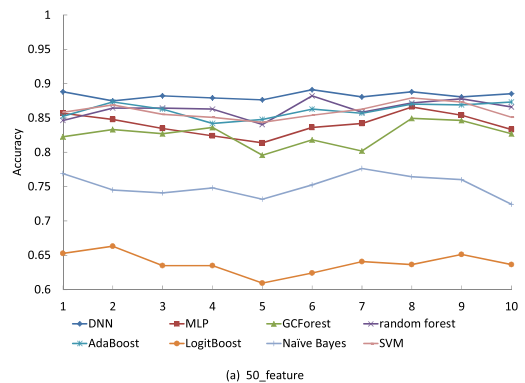
thus reducing the feature number from 50 to 44 features. Among the top 44 selected features, the feature 'H01' has the highest ranking score of 0.902, and the feature 'L01' has the lowest ranking score of 0.1665.

B. PERFORMANCE OF CLASSIFICATION MODELS
Figure 4 shows the classification performance for each of the 10 rounds individually. (a) shows the accuracy analysis results using 50 features selected by the Neuropsychologists. (b) shows the accuracy analysis results using the top

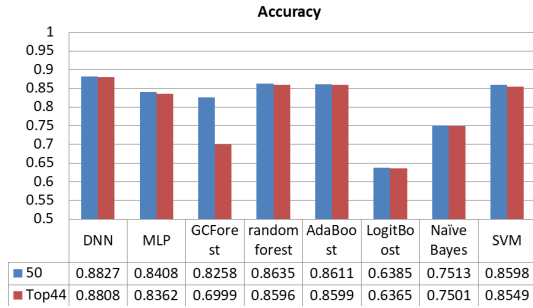**Accuracy**

	DNN	MLP	GCForest	random forest	AdaBoost	LogitBoost	Naïve Bayes	SVM
50	0.8827	0.8408	0.8258	0.8635	0.8611	0.6385	0.7513	0.8598
Top44	0.8808	0.8362	0.6999	0.8596	0.8599	0.6365	0.7501	0.8549

FIGURE 5. Comparison of the accuracy obtained by DNN and other classifiers.

F1-score(50_Feature)

	Normal	MCI	VMD	Mild	Moderate	Severe
DNN	0.89	0.89	0.74	0.85	0.88	0.92
MLP	0.84	0.82	0.57	0.86	0.89	0.91
GCForest	0.85	0.83	0.66	0.84	0.83	0.89
random forest	0.86	0.87	0.76	0.87	0.87	0.91
AdaBoost	0.84	0.85	0.69	0.88	0.9	0.94
LogitBoost	0	0.68	0.09	0.69	0.73	0.81
Naïve Bayes	0.72	0.72	0.57	0.78	0.78	0.82
SVM	0.85	0.85	0.67	0.87	0.89	0.93

(a)

F1-score(Top44_Feature)

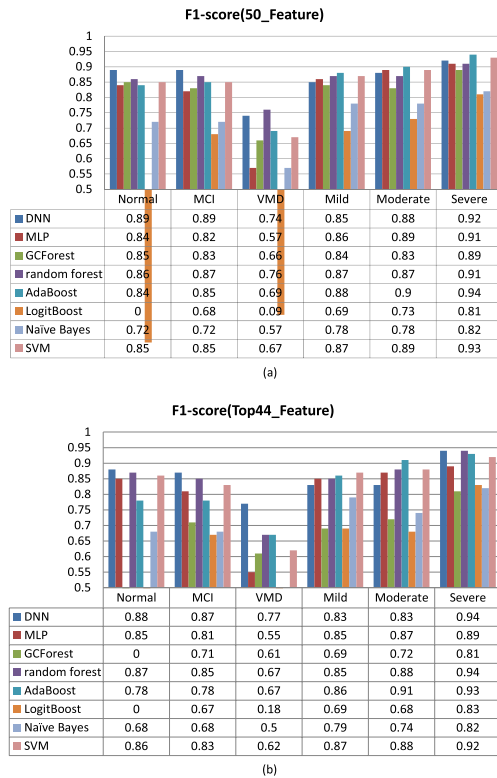|	Normal	MCI	VMD	Mild	Moderate	Severe
DNN	0.88	0.87	0.77	0.83	0.83	0.94
MLP	0.85	0.81	0.55	0.85	0.87	0.89
GCForest	0	0.71	0.61	0.69	0.72	0.81
random forest	0.87	0.85	0.67	0.85	0.88	0.94
AdaBoost	0.78	0.78	0.67	0.86	0.91	0.93
LogitBoost	0	0.67	0.18	0.69	0.68	0.83
Naïve Bayes	0.68	0.68	0.5	0.79	0.74	0.82
SVM	0.86	0.83	0.62	0.87	0.88	0.92

(b)

FIGURE 6. Performance of F1-score in the classification of normal, MCI, VMD, Mild, Moderate and Severe using classifiers.

44 features selected by information gain score. The accuracy performance of the DNN reaches a plateau in the 10 rounds, which are better than other algorithms. The performance of the model reduces when lower features are used as input into the classifier.

Figure 5 shows the average results in 10 rounds of the comparison of our DNN accuracy performance results and other well-known classifiers (MLP, GCForest, random forest, AdaBoost, LogitBoost, Naïve Bayes and SVM) for the same dataset. When using 50 features, the best accuracy was obtained by the DNN classifier (accuracy = 0.8748), followed by the MLP classifier (accuracy = 0.851). When using the top 44 features, the DNN classifier performs the best (accuracy = 0.8808), followed by the AdaBoost (accuracy = 0.8599).

C. MULTI-CLASS CLASSIFICATION

Figure 6 compares the F1-score performance of the 10 rounds in the classification of normal, MCI, VMD, Mild, Moderate and Severe using DNN, MLP, GCForest, random forest, AdaBoost, LogitBoost, Naïve Bayes and SVM. As shown in Figure 6, when using all the 50 features, the DNN algorithm effectively improved the overall performance in classifying normal (F1-score = 0.89), MCI (F1-score = 0.89), VMD (F1-score = 0.74), Mild (F1-score = 0.85), Moderate (F1-score = 0.88) and Severe (F1-score = 0.92). When using the top 44 features selected by information gain score. The DNN algorithm performed best result in screening the normal (F1-score = 0.88), MCI (F1-score = 0.87), VMD (F1-score = 0.77) and Severe (F1-score = 0.94), and poorest in Mild (F1-score = 0.83) and Moderate (F1-score = 0.83) categories.

IV. DISCUSSION

In this study, we proposed a deep neural network classification model based on the Keras framework. In order to evaluate the advantages of our proposed method, we compared two indicators, accuracy and F1-score. In addition, we compared our method with other well-known machine learning methods. The results showed our DNN method had a stable classification performance, higher classification accuracy and performed well in dealing with class imbalance problems. It has great potential for clinical application. We will discuss these in detail below.

From the perspective of model classification stability and accuracy, when using 50 features by the Neuropsychologists, the DNN model shows higher stability and the classification accuracy is the highest compared with the other seven algorithms (MLP, GCForest, random forest, AdaBoost, LogitBoost, Naive Bayes and SVM), basically stable at around 0.88. When it comes to the classification accuracy of each category, our results show that the DNN model improved the overall performance of the classification accuracy of each category.

We further studied the classification performance of the DNN model after reducing features by information gain feature selection, which is to simplify the procedure of diagnosis and enhance the practicality in clinic. The overall classification performance of the model has decreased after the reduction. In order to ensure the classification accuracy of the model, we set the threshold of information gain fraction to 0.16, thus discarding some redundant features. Compared with the six classification models, the DNN performed the best accuracy with 0.88, which also showed good results for identifying normal (F1-score = 0.88), MCI (F1-score = 0.87), VMD (F1-score = 0.77) and Severe (F1-score = 0.94).

V. CONCLUSION

We proposed a new approach to diagnosing normal, MCI, VMD, Mild, Moderate, and Severe using a deep learning approach, more specifically, a deep neural network classification model based on the Keras framework. By using the

real-world dataset, i.e., the register-based database in the Show Chwan Health System, we tested and validated our method. Overall, the results of this project show that the proposed DNN model provides a tool with accurate and stable performance for clinicians to diagnose the early stages of dementia. Our future work will be carried out from neuroimaging to further improve our diagnostic model.

REFERENCES

[1] Y. He, Z. Chen, G. Gong, and A. Evans, "Neuronal networks in Alzheimer's disease," *Neuroscientist*, vol. 15, no. 4, pp. 333–350, 2009, doi: 10.1177/1073858409334423.

[2] The Lancet Editorial, "Dementia burden coming into focus," *Lancet*, vol. 390, no. 10113, p. 2606, 2017, doi: 10.1016/S0140-6736(17)33304-4.

[3] J. Gratwicke, M. Jahanshahi, and T. Foltynie, "Parkinson's disease dementia: A neural networks perspective," *Brain*, vol. 138, no. 6, pp. 1454–1476, 2015, doi:10.1093/brain/awv104.

[4] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017, doi:10.1016/j.neuroimage.2017.03.057.

[5] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cogn. Syst. Res.*, vol. 57, pp. 147–159, Oct. 2019, doi:10.1016/j.cogsys.2018.12.015.

[6] D. Lu, K. Popuri, G. W. Ding, and R. Balachandar, "Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images," *Sci. Rep.*, vol. 8, Oct. 2018, Art. no. 5697, doi: 10.1038/s41598-018-22871-z.

[7] C. V. Dolph, M. Alam, Z. Shboul, M. D. Samad, and K. M. Iftekharuddin, "Deep learning of texture and structural features for multiclass Alzheimer's disease classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2259–2266, doi:10.1109/IJCNN.2017.7966129.

[8] S. O. Orimaye, J. S.-M. Wong, and C. P. Wong, "Deep language space neural network for classifying mild cognitive impairment and Alzheimer-type dementia," *PLoS ONE*, vol. 13, no. 11, 2018, Art. no. e0205636, doi:10.1371/journal.pone.0205636.

[9] H. S. Choi, J. Y. Choe, and H. Kim, "Deep learning based low-cost high-accuracy diagnostic framework for dementia using comprehensive neuropsychological assessment profiles," *BMC Geriatr.*, vol. 18, no. 1, 2018, Art. no. 234, doi: 10.1186/s12877-018-0915-z.

[10] C. Themistocleous, M. Eckerström, and D. Kokkinakis, "Identification of mild cognitive impairment from speech in swedish using deep sequential neural networks," *Frontiers Neurol.*, vol. 9, Nov. 2018, Art. no. 975, doi:10.3389/fneur.2018.00975.

[11] P.-Y. Chiu, H. Tang, C.-Y. Wei, C. Zhang, W. A. Hung, and G. Zhou, "NMD-12: A new machine-learning derived screening instrument to detect mild cognitive impairment and dementia," *PLoS ONE*, vol. 14, no. 3, 2019, Art. no. e0213430.

[12] P. Y. Chiu, C. Y. Wei, and G. U. Hung, "Preliminary study of the history-based artificial intelligent clinical dementia diagnostic system," *Show Chwan Med. J.*, vol. 18, no. 1, pp. 018–027, 2019, doi: 10.3966/156104972019061801003.

[13] C. M. Lin, G. U. Hung, C. Y. Wei, R. C. Tzeng, and P. Y. Chiu, "An informant-based simple questionnaire for language assessment in neurodegenerative disorders," *Dement. Geriatr. Cogn. Disord.*, vol. 46, nos. 3–4, pp. 207–216, Oct. 2018, doi: 10.1159/000493540.

[14] P. Y. Chiu, G. U. Hung, and C. Y. Wei, "History-based questionnaire for the diagnosis of severity and subtypes of dementia: Design and verify," in *Proc. Alzheimer's Assoc. Int. Conf. (AAIC)*, 2019.

[15] P. Y. Chiu and C. F. Tsai, "Freezing of speech single questionnaire as a screening tool for dementia with Lewy bodies," in *Proc. Alzheimer's Assoc. Int. Conf. (AAIC)*, 2019.

[16] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2012, doi: 10.1007/s13398-014-0173-7.2.

[17] T. Kim and T. Adali, "Complex backpropagation neural network using elementary transcendental activation functions," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2001, pp. 1281–1284.